GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 8, 2004, 10:25:26 ; Search time 48 Seconds
(without alignments)
2331.297 Million cell updates/sec

Title: US-09-907-907A-42
Perfect score: 3557
Sequence: 1 DGPFLLPRRDRALTQLQVRA.....TAVLLHNTQLDNERLNILLP 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3557	100.0	705	21	AAB08732	Amino acid sequenc
2	3498	98.3	783	23	ABP69432	Human polypeptide
3	3030	85.2	675	23	ABG30875	Human polynucleoti
4	2285	64.2	504	22	AAB92684	Human protein sequ
5	1982	55.7	899	22	ABG08546	Novel human diagno
6	1982	55.7	899	22	ABG17275	Novel human diagno
7	1917	53.9	439	22	ABG17276	Novel human diagno
8	1842.5	51.8	748	22	ABB58546	Drosophila melanog
9	1228.5	34.5	541	22	ABG08547	Novel human diagno
10	1191.5	33.5	696	22	AAU33947	Staphylococcus aur
11	1191.5	33.5	698	22	AAU36732	Staphylococcus aur
12	1165.5	32.8	706	23	ABP40063	Staphylococcus epi
13	1158	32.6	721	22	AAU38175	Salmonella typhi c
14	1156	32.5	1034	20	AAY03792	S. aureus polypept
15	1155	32.5	734	22	AAU34719	E. coli cellular p
16	1134	31.9	709	22	AAU35396	Haemophilus influe
17	1128.5	31.7	704	22	AAU35289	Enterococcus faeca
18	1127	31.7	702	22	AAU33479	Enterococcus faeca
19	1127	31.7	723	23	ABB49134	Listeria monocytog
20	1126	31.7	716	23	ABP54549	Pasteurella haemol
21	1117	31.4	711	24	ABP78218	N. gonorrhoeae ami
22	1111	31.2	706	23	AAU72928	Neisseria meningit
23	1101.5	31.0	684	21	AAB44567	Virulence gene pro
24	1101.5	31.0	684	23	ABP54519	Pasteurella multic
25	1097	30.8	613	22	AAU36162	Klebsiella pneumon
26	1097	30.8	701	22	AAU36457	Pseudomonas aerugi
27	1086	30.5	709	23	ABP30163	Streptococcus poly
28	1086	30.5	713	23	ABP25678	Streptococcus poly
29	1072	30.1	710	23	ABP25679	Streptococcus poly
30	1063.5	29.9	773	23	ABB55204	Lactococcus lactis
31	1054	29.6	737	21	AAY99614	Streptococcus pneu
32	1054	29.6	737	22	AAU38106	Streptococcus pneu
33	1054	29.6	737	24	ABU00963	S. pneumoniae type
34	1045	29.4	752	22	AAU46670	Propionibacterium
35	1044.5	29.4	913	23	ABP66191	Bifidobacterium lo
36	1014.5	28.5	753	22	AAG91911	C glutamicum prote
37	997.5	28.0	752	22	AAG81201	Mycobacterium tube
38	884	24.9	568	20	AAY35655	Protein involved i
39	831.5	23.4	688	19	AAW98709	H. pylori GHPO 758
40	831.5	23.4	688	22	AAU35823	Helicobacter pylor
41	829.5	23.3	688	22	AAU35983	Helicobacter pylor
42	710	20.0	488	21	AAB44542	Virulence gene pro
43	710	20.0	488	23	ABP54494	Pasteurella multic
44	708	19.9	164	21	AAB56873	Human prostate can
45	636	17.9	358	21	AAB44590	Virulence gene pro

ALIGNMENTS

RESULT 1
AAB08732
ID AAB08732 standard; Protein; 705 AA.
XX
AC AAB08732;
XX
DT 02-JAN-2001 (first entry)
XX
DE Amino acid sequence of a human OLD-35 polypeptide.
XX
KW OLD-35; OLD-64; OLD-137; OLD-139; OLD-142; OLD-175; cancerous phenotype;
KW cellular senescence; terminal differentiation; growth suppression;
KW aging process; type I interferon; cancer cell; tissue regeneration; ss.
XX
OS Homo sapiens.
XX
PN WO200046391-A2.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-US02920.
XX
PR 02-FEB-1999; 99US-0243277.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Fisher PB, Leszcyniecka M;
XX
DR WPI; 2000-532905/48.
DR N-PSDB; AAA64608.
XX
PT Novel isolated nucleic acid encoding an OLD-35 or OLD-64 protein useful

```
PT   in the treatment and detection of e.g. cancer and diseases involving
PT   cellular senescence -
XX
PS   Disclosure; Fig 9B; 115pp; English.
XX
CC   The specification describes OLD-35, OLD-64, OLD-137, OLD-139, OLD-142
CC   and OLD-175 proteins. The OLD nucleic acids are useful for reversing
CC   the cancerous phenotype of a cancer cell, determining if a cell is
CC   senescent, growth arrested or terminally differentiated. They are also
CC   useful for reversing the aging process in a cell and degrading specific
CC   RNAs in a cell. The genes may also be used as a diagnostic indicator of
CC   cellular senescence, terminal differentiation and/or growth suppression
CC   and as a marker to identify drugs or small molecules that will induce
CC   or inhibit cellular senescence or terminal differentiation and type I
CC   interferons. The combination of Old-35 with other interacting proteins
CC   is useful for targeting the differentiation of specific cells. Old-35
CC   can be used to selectively stabilize specific mRNAs containing adenoviral
CC   rich 3' UTRs. The OLD proteins are useful for reversing the cancerous
CC   phenotype of a cancer cell and inhibiting the growth of a cancer cell.
CC   They are also useful for regenerating tissue. The present sequence
CC   represents an OLD-35 polypeptide.
XX
SQ   Sequence   705 AA;

  Query Match            100.0%;  Score 3557;  DB 21;  Length 705;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-304;
  Matches  705;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60

Qy         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120

Qy        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180

Qy        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240

Qy        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300

Qy        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360

Qy        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420

Qy        421 GKVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GKVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMD 480

Qy        481 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 540

Qy        541 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 600

Qy        601 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 660

Qy        661 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERLNILLP 705
              |||||||||||||||||||||||||||||||||||||||||||||
Db        661 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERLNILLP 705


RESULT 2
ABP69432
ID   ABP69432 standard; Protein; 783 AA.
XX
AC   ABP69432;
XX
DT   20-JAN-2003  (first entry)
XX
DE   Human polypeptide SEQ ID NO 1479.
XX
KW   Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW   cell-proliferative disorder; neurodegenerative disease; bacterial;
KW   Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW   multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW   arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW   antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW   haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW   antiarthritic.
XX
OS   Homo sapiens.
XX
PN   WO200270539-A2.
XX
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US05095.
XX
PR   05-MAR-2001; 2001US-0799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Zhou P,  Goodrich RW,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Ma Y,  Yamazaki V,  Chen R,  Wang Z,  Ghosh M;
PI   Wehrman T,  Wang J,  Wang D,  Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   N-PSDB; ABZ11649.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT   platelet or coagulation disorders  -
XX
PS   Claim 9; SEQ ID NO 1479; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences
CC   (ABZ11119-ABZ12066) or their mature protein coding portion, active domain
CC   coding protein or complementary sequences. The polynucleotides are useful
CC   for identifying expressed genes or for physical mapping of human genome.
CC   The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC   weight markers, as a food supplement, for generating antibodies, in
CC   medical imaging, screening and diagnostic assays and for treating
CC   cell-proliferative disorders (cancer), neurodegenerative diseases
CC   (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC   sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC   disorders, platelet or coagulation disorders, wound, burns, incision,
CC   ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC   parasitic), arthritis, etc.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   783 AA;

  Query Match            98.3%;  Score 3498;  DB 23;  Length 783;
  Best Local Similarity  98.9%;  Pred. No. 3.7e-299;
  Matches  692;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 77

Qy         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db         78 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREIGTSDKEILTSRIIDRS 137

Qy        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 197

Qy        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        198 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYSQQIIQ 257

Qy        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 CIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 317

Qy        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 377

Qy        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 437

Qy        421 GKVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 GKVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMD 497

Qy        481 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 557

Qy        541 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 617

Qy        601 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 660
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        618 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSAMHEARDFITEICKDDQEQQLE 677

Qy        661 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
              |||||||||||||||||||||||||||||||||||  ::
Db        678 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDQRKI 717
```

RESULT 3
ABG30875
ID ABG30875 standard; Protein; 675 AA.
XX
AC ABG30875;
XX
DT 21-OCT-2002 (first entry)
XX
DE Human polynucleotide phosphorylase 74.25.
XX
KW Human; enzyme; polynucleotide phosphorylase 74.25; malignant tumour;
KW haemopathy; human immunodeficiency virus infection; HIV;
KW immunological disease; inflammation.
XX
OS Homo sapiens.
XX
PN CN1341720-A.
XX
PD 27-MAR-2002.
XX
PF 05-SEP-2000; 2000CN-0119892.
XX
PR 05-SEP-2000; 2000CN-0119892.
XX
PA (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.
XX
PI Mao Y, Xie Y;
XX
DR WPI; 2002-501203/54.
DR N-PSDB; ABK89228.
XX
PT New polypeptide-polynucleotide phosphorylase 74.25 for treating
PT malignant tumour, haemopathy, human immunodeficiency virus infection,
PT immunological disease and various inflammations -
XX
PS Claim 1; Page 25-26 (Disclosure); 33pp; Chinese.
XX
CC The present invention discloses a new polypeptide-polynucleotide
CC phosphorylase 74.25, a polynucleotide encoding the polypeptide and a
CC method for producing the polypeptide using DNA recombination technology.
CC The invention also discloses a method for curing several diseases, such
CC as malignant tumour, haemopathy, human immunodeficiency virus (HIV)
CC infection, immunological disease and various inflammations by using the
CC polypeptide. The invention also discloses an antagonist for resisting the
CC polypeptide and its therapeutic action, and also discloses the
CC application of the polynucleotide encoding the new polynucleotide
CC phosphorylase 74.25. The present sequence represents the
CC polynucleotide phosphorylase 74.25.
XX
SQ Sequence 675 AA;

Query Match 85.2%; Score 3030; DB 23; Length 675;
Best Local Similarity 94.5%; Pred. No. 5.8e-258;
Matches 602; Conservative 3; Mismatches 4; Indels 28; Gaps 1;

```
Qy         64 MVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRP 123
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db          1 MVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREIGTSDKEILTSRIIDRSIRP 60

Qy        124 LFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEY 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEY 120

Qy        184 VVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQ 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQ 180

Qy        244 QLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDT 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDT 240

Qy        304 EEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTL 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 EEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTL 300

Qy        364 HGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKV 423
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        301 HGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMPHYEFPPYATNEIGKV 360

Qy        424 TGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGV 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGV 420

Qy        484 PISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQAD 543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 PISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQAD 480

Qy        544 IKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKF 603
              |||||||||||||||||||||||||||||||||||||||||||||
Db        481 IKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVV------------- 527

Qy        604 VGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGA 663
                                  |||||||||||||||||| |||||||||||||||||||||
Db        528 --------------GVTISQVDEETFSVFAPTPSAMHEARDFITEICKDDQEQQLEFGA 572

Qy        664 VYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
              |||||||||||||||||||||||||||||||||||  ::
```

```
Db        573 VYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDQRKI 609

RESULT 4
AAB92684
ID   AAB92684 standard; Protein; 504 AA.
XX
AC   AAB92684;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:11065.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11065; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   504 AA;

  Query Match             64.2%;  Score 2285;  DB 22;  Length 504;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-192;
  Matches  450;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 77

Qy         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 137

Qy        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 197

Qy        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 257

Qy        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 317

Qy        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 377

Qy        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 437

Qy        421 GKVTGLNRRELGHGALAEKALYPVIPRDFP 450
              |||||||||||||||||||||||||||||
Db        438 GKVTGLNRRELGHGALAEKALYPVIPRDFP 467

RESULT 5
ABG08546
ID   ABG08546 standard; Protein; 899 AA.
XX
AC   ABG08546;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #8537.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS72733.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 38905; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
```

```
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   899 AA;

  Query Match             55.7%;  Score 1982;  DB 22;  Length 899;
  Best Local Similarity   73.2%;  Pred. No. 2.8e-165;
  Matches  408;  Conservative   18;  Mismatches   29;  Indels  102;  Gaps    2;

Qy         10 DRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVS 69
              |||| ||||||| || ||:|||:||||||||||||||||:||||:|||||||| ||||
Db        440 DRALIQLQVRALLSSTGSQAVAMDLGNRKLEISSGKLARFANGSAVIQSGDTAVMATAVS 499

Qy         70 KTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGY 129
              ||||||||||||||||||||
Db        500 KTKPSPSQFMPLVVDYRQKAAA---------------------------------------- 521

Qy        130 FYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTR 189
                                                          :|: ||| ||||||
Db        522 ---------------------------------------------VGMTDGECVVNPTR 535

Qy        190 KEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKET 249
              |||||||||:|||||||| ||||||||||||||||||||||||||||||||||||||||
Db        536 KEMSSSTLNVVVAGAPKSQTVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKEI 595

Qy        250 GVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKE 309
              |||||||||||||| ||||: ||| ||||||||||||||:||||||||||||||||||
Db        596 GVTKRTPQKLFTPSQEIVKHAHKLTMERLYAVFTDYEHDKISRDEAVNKIRLDTEEQLKE 655

Qy        310 KFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALF 369
              ||| | |||||||| |||||||||:|||||||||||||| |||:|||||||||||| ||
Db        656 IFPEVDLYEIIESFNTVAKEVFRSIILNEYKRCDGRDLTLLRNISCEVDMFKTLHGSELF 715

Qy        370 QRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEPPPYATNEIGKVTGLNRR 429
              ||||||:|| |||||||| || |:||| |||||||||||||||||||| ||||:|||
Db        716 QRGQTQLLCAVTFDSLESSIKLDRVITTINGIKDKNFMLHYEPPPYATNEIVKVTGINRR 775

Qy        430 ELGHGALAEKALYPVIPRDFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAV 489
              ||| |||||||||||||:|||| ||||||||||||||||||
Db        776 ELGPGALAEKALYPVIPKDFPPFAIRVTSEVLESNGSSSMASAC---------------- 818

Qy        490 AGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGI 549
              |||:|| |||| |||||||| |||||||| ||||||||||||||||||||||||||||||
Db        819 -GVAMGLATKTDLEKGEIEDYHLLTDILGIEAYNGDMDFKIAGTNKGITALQADIKLPGI 877

Qy        550 PIKIVMEAIQQASVAKK 566
              :|||||||||||||||:
Db        878 TMKIVMEAIQQASVAKR 894


RESULT 6
ABG17275
ID   ABG17275 standard; Protein; 899 AA.
XX
AC   ABG17275;
XX
DT   18-FEB-2002  (first entry)

XX
DE   Novel human diagnostic protein #17266.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS81462.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 20; SEQ ID No 47634; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   899 AA;

  Query Match             55.7%;  Score 1982;  DB 22;  Length 899;
  Best Local Similarity   73.2%;  Pred. No. 2.8e-165;
  Matches  408;  Conservative   18;  Mismatches   29;  Indels  102;  Gaps    2;

Qy         10 DRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVS 69
              |||| ||||||| || ||:|||:||||||||||||||||:||||:|||||||| ||||
Db        440 DRALIQLQVRALLSSTGSQAVAMDLGNRKLEISSGKLARFANGSAVIQSGDTAVMATAVS 499

Qy         70 KTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGY 129
              ||||||||||||||||||||
Db        500 KTKPSPSQFMPLVVDYRQKAAA---------------------------------------- 521

Qy        130 FYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTR 189
                                                          :|: ||| ||||||
Db        522 ---------------------------------------------VGMTDGECVVNPTR 535

Qy        190 KEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKET 249
              |||||||||:|||||||| ||||||||||||||||||||||||||||||||||||||||
```

```
Db        536 KEMSSSTLNVVVAGAPKSQTVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKEI 595

Qy        250 GVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKE 309
              |||||||||||||| |||| ||| ||||||||||||||:|||||||||||||||||||
Db        596 GVTKRTPQKLFTPSQEIVKHAHKLTMERLYAVFTDYEHDKISRDEAVNKIRLDTEEQLKE 655

Qy        310 KFPEADPYEIIESFNVVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALF 369
              ||| | |||||||| |||||||||:|||||||||||||| |||:|||||||||||||| ||
Db        656 IFPEVDLYEIIESFNTVAKEVFRSIILNEYKRCDGRDLTLLRNISCEVDMFKTLHGSELF 715

Qy        370 QRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRR 429
              ||||||:|| |||||||| || |:||| ||||||||||||||||||||| ||||:|||
Db        716 QRGQTQLLCAVTFDSLESSIKLDRVITTINGIKDKNFMLHYEFPPYATNEIVKVTGINRR 775

Qy        430 ELGHGALAEKALYPVIPRDFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAV 489
              ||| |||||||||||||:|||| ||||||||||||||||||
Db        776 ELGPGALAEKALYPVIPKDFPPFAIRVTSEVLESNGSSSMASAC---------------- 818

Qy        490 AGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGI 549
              |||:|| |||| |||||||| |||||||| ||||||||||||||||||||||||||||||
Db        819 -GVAMGLATKTDLEKGEIEDYHLLTDILGIEAYNGDMDFKIAGTNKGITALQADIKLPGI 877

Qy        550 PIKIVMEAIQQASVAKK 566
              :|||||||||||||||:
Db        878 TMKIVMEAIQQASVAKR 894
```

RESULT 7
ABG17276
ID ABG17276 standard; Protein; 439 AA.
XX
AC ABG17276;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17267.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS81463.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 47635; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 439 AA;

```
  Query Match             53.9%;  Score 1917;  DB 22;  Length 439;
  Best Local Similarity   78.7%;  Pred. No. 4.6e-160;
  Matches  396;  Conservative    6;  Mismatches   15;  Indels   86;  Gaps    3;

Qy          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 77

Qy         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
              | |||||||||||||| ||||||||||||
Db         78 TTVMVTAVSKTKPSPSQIMPLVVDYRQKAAA---------------------------- 108

Qy        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180
                                                                ||||||||
Db        109 ------------------------------------------------------AVRIGIID 116

Qy        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        117 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 176

Qy        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 236

Qy        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 296

Qy        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        297 KTLHGSALFQRGQTQVLCTATFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 356

Qy        421 GKVTGLNRRELGHGALAEKALYPVIPRDFPPFTIRVTSEVLESNGSSSMASACGGSLALMD 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||    ::
Db        357 GKVTGLNRRELGHGALAEKALYPVIPRDFPPFTIRVTSEVLESNGSSSMASACWRKFSI-- 414

Qy        481 SGVPISSAVA---GVAIGLVTKT 500
              :|   |: :       :||||||
Db        415 NGFRGSNFICCCRPYKLGLVTKT 437
```

RESULT 8
ABB58546
ID ABB58546 standard; Protein; 748 AA.
XX
AC ABB58546;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2430.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.

XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL02649.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 2430; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 748 AA;

Query Match 51.8%; Score 1842.5; DB 22; Length 748;
Best Local Similarity 52.8%; Pred. No. 4.1e-153;
Matches 364; Conservative 124; Mismatches 170; Indels 31; Gaps 6;

```
Qy         19 RALWSSAGSRAVAVDL---GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSP 75
Db         29 RGIQSSSNGEAPSVEVNFSNGRNMTFSSGRLARFANGTAVCQMGDTAVMVTAVAKAKPNP 88

Qy         76 SQ-FMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQ 134
Db         89 GQGFMPLVVDYRLKNAASGRIPMNFMRRELGPSEKEILSARLIDRSLRPLFHKDYRTETQ 148

Qy        135 VLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSS 194
Db        149 LVCNMLAMDAVHSPDVLAINAASMALSLSDIPWNGPIGAVRVGLCDGEVLINPTRRELQT 208

Qy        195 STLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKR 254
Db        209 SQLDLVVSATKQNLVVMLEGKGNVVLQQDLLKAIKQGTREAQFIIHEIERLQKAYGRQKR 268

Qy        255 TPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEA 314
Db        269 EVEVAAEVDPELGKAVRSMCEMRLREIFQDSTHDKMSRDNAVNEVRSNVIDKVWSSFPDT 328

Qy        315 DPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQT 374
Db        329 EPSLITEQFNQTSRTIFRELIFERGLRCDGRDYDQLRNISCQVDMYKPLHGSALFQRGQT 388

Qy        375 QVLCTVTFDSLESGIKSDQVITAIN--GIKDKNFMLHYEFPPYATNEIGKVTGLNRRELG 432
Db        389 QVFCTVSLDSQESAMKLDS-LAALDSGGLKAKNFMLHYEFPPYATGEVGRIGPVGRREMG 447

Qy        433 HGALAEKALYPVIPRDFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGV 492
Db        448 HGALAERSLLPTLPNDYPFTVRLTSEVLESNGSSSMASVCGGSLALMDAGVPVSAPAAGV 507

Qy        493 AIGLVTK-TDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPI 551
Db        508 AIGLVTKFENDDTKHLQDYRILTDILGIEDYMGDMD----------------------M 544

Qy        552 KIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNL 611
Db        545 KVVMESLQKATDAKSNILDIMSEAIREPRKYPKESWPVSETLTVEPQQRAQLIGPSGLHM 604

Qy        612 KKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITE 671
Db        605 KRIYLETGTSLTAVDETHFNVFAPSQAAMDEAKELIEGYMVKERVPDLEFGGIYTAKITE 664

Qy        672 IRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
Db        665 LRDTGVMVILYPSMPPALLHNSQLDQRKI 693
```

RESULT 9
ABG08547
ID ABG08547 standard; Protein; 541 AA.
XX
AC ABG08547;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #8538.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS72734.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 38906; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO

```
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   541 AA;

  Query Match           34.5%;  Score 1228.5;  DB 22;  Length 541;
  Best Local Similarity   51.7%;  Pred. No. 3.7e-99;
  Matches  306;  Conservative   48;  Mismatches  133;  Indels  105;  Gaps   18;

Qy         97 TNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGA 156
              || : :  | : :|:    : :::: |: |  : :  :|   :|  :|:
Db          6 TNEIGKVTGLNRRELGHGALAEKALYPVIPRDFPPFTIRVTSEVLESNGMR---------- 55

Qy        157 SVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAP------KSQIV 210
                          |||           :   ||  :  | |  |      :: :
Db         56 ------------------RIG----------QPPASSRPVRRVRAPLPPPLLEVRAPLP 86

Qy        211 MLEASAENILQQDFCHAIKVGVKYTQQI--IQGIQQLVKETGVTKRTPQKLFTPSPEIVK 268
                | | :    |  ::: |:|  | | | ||:               | |
Db         87 ASRPSGEVLCTVTF-DSLESGIKSDQVITAINGIK------------DKNF-------- 124

Qy        269 YTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLD----TEEQLKEKFPEADPYEIIESFN 324
                     |:  |  |  ::: :   :|:  |       |: |     |   |: |
Db        125 --------MLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIPRDFPFTI----- 171

Qy        325 VVAKEVFRSIVLNEYKRCDGRDLTS--LRNVSCE-----VDMFKTLHGSALFQRGQTQVL 377
                |  ||  |   |  :|      :|  :| |        :::   |  |    |   :||
Db        172 RVTSEVLES---NGMRRIGQPPASSRPVRRVRAPLPPPLLEVRAPLPAS----RPSGEVL 224

Qy        378 CTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALA 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 CTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALA 284

Qy        438 EKALYPVIPRDFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAG-VAIGL 496
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        285 EKALYPVIPRDFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGRTNXDW 344

Qy        497 VTKTDPEKGEIEDYRLLTDI-LGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVM 555
                | |||||||    :   | ||||||||||||||||||||||||||||||||||||||
Db        345 SPKPDPEKGEIRKIIVCXQIFLGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVM 404

Qy        556 EAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQ 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 EAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQ 464

Qy        616 AETGVTISQVDEET-FSVFAPTPSVMHEARDF-ITEICKDDQEQQL-EFGAV 664
              |||||||||||||| |      | :| | :    |||||||||| :    |||
Db        465 AETGVTISQVDEETVFCNLHQHPVLMLEGKRLSFTEICKDDQEQAIRNLGAV 516


RESULT 10
AAU33947
ID   AAU33947 standard; Protein; 696 AA.
XX
AC   AAU33947;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #223.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS51806.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 5443; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   696 AA;

  Query Match           33.5%;  Score 1191.5;  DB 22;  Length 696;
  Best Local Similarity   39.2%;  Pred. No. 1e-95;
  Matches  274;  Conservative  117;  Mismatches  249;  Indels   59;  Gaps   12;

Qy         24 SAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVV 83
              |  :    :   | | | :|:||: |:|: :|: ||| |: || :  :|    | || |
Db          2 SQEKKVFKTEWAGRSLTIETGQLAKQANGAVLVRYGDTVVLSTATASKEPRDGDFFPLTV 61

Qy         84 DYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVD 143
              :| :|  |||:||  : :||     |    ||:|:||| |||||| || :| |::  :|: |
Db         62 NYEEKMYAAGKIPGGFKKREGRPGDDATLTARLIDRPIRPLFPKGYKHDVQIMNMVLSAD 121

Qy        144 GVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAG 203
                 | : |: |:|:|||:|||||: ||:  | :| |||:|::||| :|    | |:| |||
Db        122 PDCSPQMAAMIGSSMALSVSDIPFQGPIAGVNVGYIDGKYIINPTVEEKEVSRLDLEVAG 181

Qy        204 APKSQIVMLEASAENILQQDFCHAIKVG-------VKYTQQIIQGIQQLVKETGVTKRTP 256
                  | : |:|| |  | :|:   ||  |       | : |||:  || : :|
Db        182 -HKDAVNMVEAGASEITEQEMLEAIFFGHEEIQRLVDFQQQIVDHIQPVKQE-------- 232

Qy        257 QKLFTPS-------PEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKE 309
                 | |:          |   | :  ::      |     ||  ||| :: ::   || : |
Db        233 ---FIPAERDEALVERIKSLTEEKGLKETVLTF-----DKQQRDENLDNLK---EEIVNE 281

Qy        310 KFPEADP------YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTL 363
                | ||       |:    | : ||  | :: :|  | |||    :| :  || :
Db        282 FIDEEDPENELLIKEVYAILNELVKEEVRRLIADEKIRPDGRKPDEIRPLDSEVGILPRT 341

Qy        364 HGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKV 423
              ||| || ||||| |  :|  :|       |:|  :   ::| || || || ::  | | |
Db        342 HGSGLFTRGQTQALSVLTLGAL----GDYQLIDGLGPEEEKRFMHHYNFPNFSVGETGPV 397

Qy        424 TGLNRRELGHGALAEKALYPVIP--RDFPPFTIRVTSEVLESNGSSSMASACGGSLALMDS 481
```

```
                  |||:|||||  |:||  :||   |||:|||: ||||||||||| || || :|||||:
Db        398 RAPGRREIGHGALGERALKYIIPDTADFPYTIRIVSEVLESNGSSSQASICGSTLALMDA 457

Qy        482 GVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQ 541
              ||||  : |||:|:||||: |        | :|||| |:||  ||||||:||| :||||:|
Db        458 GVPIKAPVAGIAMGLVTRED-------SYTILTDIQGMEDALGDMDFKVAGTKEGITAIQ 510

Qy        542 ADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRA 601
              |||: |:  :|: ||:||   : ||:  | :|| :||       | |:: :   |
Db        511 MDIKIDGLTREIIEEALEQARRGRLEIMNHMLQTIDQPRTELSAYAPKVVTMTIKPDKIR 570

Qy        602 KFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEF 661
              :|||  : ::  |||| :   : |  : |   :::: ||: | || :     : |
Db        571 DVIGPGGKKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAREIIEEITR-----EAEV 625

Qy        662 GAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
              |  | ||:  |   |  | |:|   | ||| :|:   |:
Db        626 GQTYQATVKRIEKYGAFVGLFPGKDA-LLHISQISKNRI 663


RESULT 11
AAU36732
ID   AAU36732 standard; Protein; 698 AA.
XX
AC   AAU36732;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #902.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54591.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12325; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   698 AA;

  Query Match             33.5%;  Score 1191.5;  DB 22;  Length 698;
  Best Local Similarity   39.2%;  Pred. No. 1e-95;
  Matches  274;  Conservative  117;  Mismatches  249;  Indels   59;  Gaps   12;

Qy         24 SAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVV 83
              |   :    :   | | | :|:||: |:|: :|: ||| |: || :  :|    | || |
Db          2 SQEKKVFKTEWAGRSLTIETGQLAKQANGAVLVRYGDTVVLSTATASKEPRDGDFFPLTV 61

Qy         84 DYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVD 143
              :| :|  |||:||  : :||    |   ||:|:||| |||||| || :|  |::  :|: |
Db         62 NYEEKMYAAGKIPGGFKKREGRPGDDATLTARLIDRPIRPLFPKGYKHDVQIMNMVLSAD 121

Qy        144 GVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAG 203
                | : |: |:|:|||:||||: ||:  | :| |||:|::||| :|   | |:| |||
Db        122 PDCSPQMAAMIGSSMALSVSDIPFQGPIAGVNVGYIDGKYIINPTVEEKEVSRLDLEVAG 181

Qy        204 APKSQIVMLEASAENILQQDFCHAIKVG-------VKYTQQIIQGIQQLVKETGVTKRTP 256
              |  : |:|| |  | :|:   || |         | : |||:  || : :|
Db        182 -HKDAVNMVEAGASEITEQEMLEAIFFGHEEIQRLVDFQQQIVDHIQPVKQE-------- 232

Qy        257 QKLFTPS-------PEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKE 309
              | |:         |   | :  ::      |      ||  ||| :: ::   || : |
Db        233 ---FIPAERDEALVERIKSLTEEKGLKETVLTF-----DKQQRDENLDNLK---EEIVNE 281

Qy        310 KFPEADP------YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTL 363
              | ||        |:    | : ||  | :: :|  | |||    :| :  || :
Db        282 FIDEEDPENELLIKEVYAILNELVKEEVRRLIADEKIRPDGRKPDEIRPLDSEVGILPRT 341

Qy        364 HGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKV 423
              ||| || ||||| |  :|  :|       |:|  :   ::| || || || ::  | | |
Db        342 HGSGLFTRGQTQALSVLTLGAL----GDYQLIDGLGPEEEKRFMHHYNFPNFSVGETGPV 397

Qy        424 TGLNRRELGHGALAEKALYPVIP--RDFPPFTIRVTSEVLESNGSSSMASACGGSLALMDS 481
              |||:||||| |:||  :||   |||:|||: |||||||||| || || :|||||:
Db        398 RAPGRREIGHGALGERALKYIIPDTADFPYTIRIVSEVLESNGSSSQASICGSTLALMDA 457

Qy        482 GVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQ 541
              |||| : |||:|:||||: |        | :|||| |:||  ||||||:||| :||||:|
Db        458 GVPIKAPVAGIAMGLVTRED-------SYTILTDIQGMEDALGDMDFKVAGTKEGITAIQ 510

Qy        542 ADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRA 601
              |||: |:  :|: ||:||   : ||:  | :|| :||       | |:: :   |
Db        511 MDIKIDGLTREIIEEALEQARRGRLEIMNHMLQTIDQPRTELSAYAPKVVTMTIKPDKIR 570

Qy        602 KFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEF 661
              :|||  : ::  |||| :   : |  : |   :::: ||: | || :     : |
Db        571 DVIGPGGKKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAREIIEEITR-----EAEV 625

Qy        662 GAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
              |  | ||:  |   |  | |:|   | ||| :|:   |:
Db        626 GQTYQATVKRIEKYGAFVGLFPGKDA-LLHISQISKNRI 663


RESULT 12
ABP40063
ID   ABP40063 standard; Protein; 706 AA.
XX
AC   ABP40063;
XX
DT   24-JUL-2002  (first entry)
XX
DE   Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4908.
```

```
XX
KW   Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW   antibacterial; gene therapy.
XX
OS   Staphylococcus epidermidis.
XX
PN   US6380370-B1.
XX
PD   30-APR-2002.
XX
PF   13-AUG-1998;   98US-0134001.
XX
PR   14-AUG-1997;   97US-055779P.
PR   08-NOV-1997;   97US-064964P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Doucette-Stamm LA,  Bush D;
XX
DR   WPI; 2002-381255/41.
DR   N-PSDB; ABN92608.
XX
PT   Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT   polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS   Disclosure; SEQ ID 4908; 267pp; English.
XX
CC   ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC   frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC   given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC   antibacterial activity and can be used in gene therapy. The sequences
CC   can also be used in the diagnosis and treatment of bacterial infections,
CC   particularly S. epidermidis infections. The sequences can be used to
CC   screen for compounds able to interfere with the S. epidermidis life
CC   cycle or inhibit S. epidermidis infection.
CC   N.B. The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   USPTO web site.
XX
SQ   Sequence   706 AA;

  Query Match             32.8%;  Score 1165.5;  DB 23;  Length 706;
  Best Local Similarity   38.7%;  Pred. No. 2.1e-93;
  Matches  270;  Conservative  123;  Mismatches  250;  Indels   55;  Gaps   13;

Qy         24 SAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVV 83
              |   :    :   | | | :|:||: |:|: :|: ||| |: || :  :|    | || |
Db          7 SQEKKVFKTEWAGRSLTIETGQLAKQANGAVLVRYGDTVVLSTATASKEPRDGDFFPLTV 66

Qy         84 DYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVD 143
              :| :|  |||:||  : :||    |:  ||:|:||| |||||| || :| |::  :|: |
Db         67 NYEEKMYAAGKIPGGPFKKREGRPGDEATLTARLIDRPIRPLFPKGYRHDVQIMNIVLSAD 126

Qy        144 GVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAG 203
               |:: |: |:|:|||:||||: ||:  | :| |||:||:||:  :   | |:| |||
Db        127 PDCSPEMAAMIGSSMALSVSDIPFQGPIAGVNVGYIDGKYVINPSVADKEISRLDLEVAG 186

Qy        204 APKSQIVMLEASAENILQQDFCHAIKVG-------VKYTQQIIQGIQQLVKETGVTKRTP 256
                |  : |:|| |  | : :   ||  |       | : |:||  || : :|
Db        187 -HKDAVNMVEAGASEITESEMLEAIFFGHEEIKRLVAFQQEIIDHIQPIKQE-------- 237

Qy        257 QKLFTP---SPEIVKYTHKLAMER--LYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKF 311
                | |     ::|:    |  ::     | |    ||  ||| :: ::   || :
Db        238 ---FVPVERDEDLVEKVKSLTEDKGLKDTVLT---FDKQQRDENLDALK---EEVVGHFL 288

Qy        312 PEADP------YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHG 365
              | ||       |:    | : ||  | :: :|  | ||| :  :| :  || :     ||
Db        289 DEEDPENETLVKEVYAILNDLIKEEVRRLIADEKIRPDGRKVDEIRPLESEVGLLPRAHG 348

Qy        366 SALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTG 425
              | || ||||| |  :|  :|       |:|  :    :| || || || ::  | | |
Db        349 SGLFTRGQTQALSVLTLGAL----GDYQLIDGLGPEVEKRFMHHYNFPNFSVGETGPVRA 404

Qy        426 LNRRELGHGALAEKALYPVIP--RDFPPTIRVTSEVLESNGSSSMASACGGSLALMDSGV 483
              |||:||||| |:||  :||  :|||:|||: |||||||||| || || :|||||:||
Db        405 PGRREIGHGALGERALRYIIPDTQDFPYTIRIVSEVLESNGSSSQASICGSTLALMDAGV 464

Qy        484 PISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQAD 543
              || : |||:|:|||| :|        | :|||| |:||  ||||||:|||  ||||:| |
Db        465 PIKAPVAGIAMGLVTRDD-------SYTILTDIQGMEDALGDMDFKVAGTKDGITAIQMD 517

Qy        544 IKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKF 603
              ||: |:  ::: ||::||   :  |:  |  || :||       | | |: :   |
Db        518 IKIDGLTREVIEEALEQARQGRLAIMDHMLHTIEQPREELSAYAPKVVTMSINPDKIRDV 577

Qy        604 VGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGA 663
              :|||| : ::  |||| :    : |  : |   :::: |:: | :| :     : | |
Db        578 IGPGGKKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAKEIIEDITR-----EAEVGQ 632

Qy        664 VYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERLN 701
              || | :  |   |  |:|:|   | ||| :|:  ||:|
Db        633 VYHAKVKRIEKYGAFVELFPGKDA-LLHISQISQERIN 669


RESULT 13
AAU38175
ID   AAU38175 standard; Protein; 721 AA.
XX
AC   AAU38175;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Salmonella typhi cellular proliferation protein #66.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Salmonella typhi.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS56034.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 13768; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
```

```
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   721 AA;

  Query Match             32.6%;  Score 1158;  DB 22;  Length 721;
  Best Local Similarity   39.9%;  Pred. No. 9.8e-93;
  Matches  270;  Conservative  117;  Mismatches  250;  Indels   40;  Gaps   11;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
Db         22 GQHTXTLETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQDFFPLTVNYQERTYAAG 81

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
Db         82 RIPGSFXRREGRPSEGETLIARLIDRPVRPLFPEGFVNEVQVIATVVSVNPQVNPDIVAM 141

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
Db        142 IGASAALSLSGIPFNGPIGAARVGYINDQYVLNPTQDELKESKLDLVVAGT-EAAVLMVE 200

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVK----- 268
Db        201 SEAELLSEDTMLGAVVFGHEQQQVVIQAINDLVKEAGKPRWDWQ------PEAVNDALNA 254

Qy        269 YTHKLAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNV 325
Db        255 RVAALAESRLSDAYRI-----TDKQERYAQVDVIKSETIEQLIAEDETLDANELGEILHA 309

Qy        326 VAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSL 385
Db        310 IEKNVVRSRVLAGEPRIDGREKDMIRGLDVRTGVLPRTHGSALFTRGETQALVTATLGT- 368

Qy        386 ESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVI 445
Db        369 ---ARDAQVLDELMGERTDSFLFHYNFPPYSVGETGMVGSPKRREIGHGRLAKRGVLAVM 425

Qy        446 P--RDFPPTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPE 503
Db        426 PDMDKFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVKEGD-- 483

Qy        504 KGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASV 563
Db        484 -----NYVVLSDILGDEDHLGDMDFKVAGSRDGISALQMDIKIEGITKEIMQVALNQAKG 538

Qy        564 AKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTIS 623
Db        539 ARLHILGVMEQAINAPRGDISEFAPRIHTIKISTDKIKDVIGKGGSVIRALTEETGTTIE 598

Qy        624 QVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYP 683
Db        599 IEDDGTVKIAATDGEKAKYAIRRIEEI-----TAEIEVGRIYNSKVTRIVDFGAFVAIGG 653

Qy        684 NMTAVLLHNTQLDNERL 700
Db        654 GKEG-LVHISQIADKRV 669


RESULT 14
AAY03792
ID   AAY03792 standard; Protein; 1034 AA.
XX
AC   AAY03792;
XX
DT   11-JUN-1999  (first entry)
XX
DE   S. aureus polypeptide.
XX
KW   Staphylococcus aureus polypeptide; thyroiditis; infective carditis;
KW   lung abscess; secretory diarrhoea; cerebral abscess; conjunctivitis;
KW   toxic shock syndrome; folliculitis; septic arthritis; antibacterial;
KW   H. pylori infection; gastric ulcer; adenocarcinoma.
XX
OS   Staphylococcus aureus.
XX
PN   EP905243-A2.
XX
PD   31-MAR-1999.
XX
PF   03-AUG-1998;   98EP-0306185.
XX
PR   05-AUG-1997;   97US-0055387.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Burnham MKR,  Lonetto MA,  Warren PV;
XX
DR   WPI; 1999-192667/17.
DR   N-PSDB; AAX31862.
XX
PT   New essential polypeptides from Staphylococcus aureus useful for
PT   treating diseases such as infective endocarditis and toxic shock
PT   syndrome
XX
PS   Claim 31; Page 50-55; 70pp; English.
XX
CC   The invention provides new Staphylococcus aureus polypeptides
CC   (AAY03781-94) and the genes (AAX31851-864) encoding them. Host cells
CC   containing vectors comprising the nucleic acid sequences are used for the
CC   recombinant expression of the proteins. The polypeptides can be used to
CC   screen for modulators for use in antibacterial therapy. The polypeptides,
CC   their antagonists and agonists are used to prevent or treat diseases
CC   caused by S. aureus such as thyroiditis, lung abscesses, infective
CC   carditis, secretory diarrhoea, cerebral abscesses, conjunctivitis, toxic
CC   shock syndrome folliculitis and septic arthritis. Screening for the
CC   presence of the polypeptides may be used to diagnose, predict the
CC   susceptibility to, or stage the progress of these S. aureus diseases and
CC   diseases caused by Helicobacter pylori such as gastric ulcers and gastric
CC   adenocarcinoma. There is not much information known about the essential
CC   genes expressed by S. aureus during infection but these new polypeptides
CC   have been identified as essential. They can therefore be used to develop
CC   antibacterial compounds specific for those essential genes and this
CC   ensures the effectiveness of the compounds in killing S. aureus. In
CC   addition, these polypeptides can be used to effectively diagnose and
CC   treat infections and diseases caused by S. aureus without the risk of
CC   development of antibiotic resistance. The present sequence represents a
CC   S. aureus polypeptide which has homology to a polyribonucleotide
CC   nucleotidyltransferase.
XX
SQ   Sequence   1034 AA;

  Query Match             32.5%;  Score 1156;  DB 20;  Length 1034;
  Best Local Similarity   39.3%;  Pred. No. 2.7e-92;
  Matches  265;  Conservative  113;  Mismatches  239;  Indels   58;  Gaps   11;

Qy         24 SAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVV 83
Db          2 SQEKKVFKTEWAGRSLTIETGQLAKQANGAVLVRYGDTVVLSTATASKEPRDGDFFPLTV 61

Qy         84 DYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVD 143
Db         62 NYEEKMYAAGKIPGGFKKREGRPGDDATLTARLIDRPIRPLFPKGYKHDVQIMNMVLSAD 121

Qy        144 GVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAG 203
Db        122 PDCSPQMAAMIGSSMALSVSDIPFQGPIAGVNVGYIDGKYIINPTVEEKEVSRLDLEVAG 181

Qy        204 APKSQIVMLEASAENILQQDFCHAIKVG-------VKYTQQIIQGIQQLVKETGVTKRTP 256
```

```
                 |  : |:|| |  | :|:   ||  |       | : |||:  || : :|
Db        182 -HKDAVNMVEAGASEITEQEMLEAIFFGHEEIQRLVDFQQQIVDHIQPVKQE-------- 232

Qy        257 QKLFTPS-------PEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKE 309
                 | |:          |   | :  ::      |     ||  ||| :: ::   || : |
Db        233 ---FIPAERDEALVERIKSLTEEKGLKETVLTF-----DKQQRDENLDNLK---EEIVNE 281

Qy        310 KFPEADP------YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTL 363
                 | ||       |:    | : ||  | :: :|  | |||    :| :  || :
Db        282 FIDEEDPENELLIKEVYAILNELVKEEVRRLIADEKIRPDGRKPDEIRPLDSEVGILPRT 341

Qy        364 HGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKV 423
               ||| || ||||| |  :|  :|       |:|  :   ::| || || || ::  | | |
Db        342 HGSGLFTRGQTQALSVLTLGAL----GDYQLIDGLGPEEEKRFMHHYNFPNFSVGETGPV 397

Qy        424 TGLNRRELGHGALAEKALYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDS 481
                  |||:|||| |:||  :||   |||:|||: ||||||||||| || || :|||||:
Db        398 RAPGRREIGHGALGERALKYIIPDTADFPYTIRIVSEVLESNGSSSQASICGSTLALMDA 457

Qy        482 GVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQ 541
               |||| : |||:|:|||||: |         | :|||| |:||  ||||||:||| :||||:|
Db        458 GVPIKAPVAGIAMGLVTRED-------SYTILTDIQGMEDALGDMDFKVAGTKEGITAIQ 510

Qy        542 ADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRA 601
                 |||: |:  :|: ||::||   : ||:  | :|| :||       | | |: :   |
Db        511 MDIKIDGLTREIIEEALEQARRGRLEIMNHMLQTIDQPRTELSAYAPKVVTMTIKPDKIR 570

Qy        602 KFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEF 661
                :|||  : ::  |||| :   : |  : |   :::: ||: | || :     : |
Db        571 DVIGPGGKKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAREIIEEITR-----EAEV 625

Qy        662 GAVYTATITEIRDTG 676
               |  | ||:  |   |
Db        626 GQTYQATVKRIEKYG 640
```

RESULT 15
AAU34719
ID AAU34719 standard; Protein; 734 AA.
XX
AC AAU34719;
XX
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #300.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52578.
XX

PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 10312; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 734 AA;

Query Match 32.5%; Score 1155; DB 22; Length 734;
Best Local Similarity 39.9%; Pred. No. 1.9e-92;
Matches 269; Conservative 118; Mismatches 252; Indels 36; Gaps 11;

```
Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
                |   : : :| :|| |  : :|    |||| || | : |  | | | || |:|:::  |||
Db         35 GQHTVTLETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQDFFPLTVNYQERTYAAG 94

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
               ||| :: |||    |: | | :|:||| |||||| |:  : ||:  :::|:    ||::|:
Db         95 RIPGSPFRREGRPSEGETLIARLIDRPIRPLFPEGFVNEVQVIATVVSVNPQVNPDIVAM 154

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
                ||| ||||| ||:|||:|| |:| |: :||:|||: |:  | |:|||||  :: ::|:|
Db        155 IGASAALSLSGIPFNGPIGAARVGYINDQYVLNPTQDELKESKLDLVVAGT-EAAVLMVE 213

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSP---EIVKYT 270
                : |: : :     |:  | :  | :|| | :|||| |     |:  : | |    :
Db        214 SEAQLLSEDQMLGAVVFGHEQQQVVIQNINELVKEAG----KPRWDWQPEPVNEALNARV 269

Qy        271 HKLAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVA 327
                ||  ||   | :      ||  |    |:|: :|  |  :     |  |: | : :
Db        270 AALAEARLSDAYRI-----TDKQERYAQVDVIKSETIATLLAEDETLDENELGEILHAIE 324

Qy        328 KEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLES 387
               | | || ||    | |||:   :| :     :    ||||| ||:|| | | |  :
Db        325 KNVVRSRVLAGEPRIDGREKDMIRGLDVRTGVLPRTHGSALFTRGETQALVTATLGT--- 381

Qy        388 GIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP- 446
                 :  ||:  : | :   |: || ||||:  | | |    |||:||| ||:: :  |:|
Db        382 -ARDAQVLDELMGERTDTFLFHYNFPPYSVGETGMVGSPKRREIGHGRLAKRGVLAVMPD 440

Qy        447 -RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKG 505
                 ||:|:|| ||: |||||||||| || ||||||:|||| :||||:|:||| : |
Db        441 MDKFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVKEGD---- 496

Qy        506 EIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAK 565
                  :| :|:|||| ||: ||||||:||:  ||:||| |||: ||  :|:  |: ||  |:
Db        497 ---NYVVLSDILGDEDHLGDMDFKVAGSRDGISALQMDIKIEGITKEIMQVALNQAKGAR 553

Qy        566 KEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQV 625
                 || :| : |: ||    |  | : |:::   |    :| ||  :: |  ||| ||
Db        554 LHILGVMEQAINAPRGDISEFAPRIHTIKINPDKIKDVIGKGGSVIRALTEETGTTIEIE 613

Qy        626 DEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNM 685
               |: |  : |        |   | ||       ::| | |||  :| | | |  | :
```

```
Db        614 DDGTVKIAATDGEKAKHAIRRIEEI-----TAEIEVGRVYTGKVTRIVDFGAFVAIGGGK 668

Qy        686 TAVLLHNTQLDNERL 700
                |:| ::|: ::|:
Db        669 EG-LVHISQIADKRV 682
```

Search completed: January 8, 2004, 10:38:47
Job time : 51 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 8, 2004, 10:37:57 ; Search time 22 Seconds
(without alignments)
1355.871 Million cell updates/sec

Title: US-09-907-907A-42
Perfect score: 3557
Sequence: 1 DGPFLLPRRDRALTQLQVRA.....TAVLLHNTQLDNERLNILLP 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1165.5	32.8	706	4	US-09-134-001C-4908	Sequence 4908, Ap
2	1107	31.1	750	4	US-09-107-532A-5868	Sequence 5868, Ap
3	1065	29.9	705	4	US-09-328-352-8112	Sequence 8112, Ap
4	976.5	27.5	697	4	US-09-252-991A-27283	Sequence 27283, A
5	884	24.9	568	4	US-09-198-452A-1073	Sequence 1073, Ap
6	226	6.4	127	4	US-09-198-452A-1074	Sequence 1074, Ap
7	140.5	3.9	1416	4	US-09-071-035-404	Sequence 404, App
8	140.5	3.9	1448	4	US-09-071-035-402	Sequence 402, App
9	140	3.9	252	4	US-09-252-991A-23776	Sequence 23776, A
10	137	3.9	191	4	US-09-252-991A-27186	Sequence 27186, A
11	128.5	3.6	244	4	US-09-328-352-5628	Sequence 5628, Ap
12	127.5	3.6	871	4	US-09-134-001C-3979	Sequence 3979, Ap
13	124.5	3.5	471	4	US-09-107-532A-5705	Sequence 5705, Ap
14	124.5	3.5	1306	3	US-08-999-774A-13	Sequence 13, Appl
15	116	3.3	1780	1	US-08-769-309A-5	Sequence 5, Appli
16	116	3.3	1780	3	US-08-994-570-5	Sequence 5, Appli
17	115	3.2	875	4	US-09-107-532A-4537	Sequence 4537, Ap
18	114	3.2	648	4	US-09-198-452A-594	Sequence 594, App
19	114	3.2	2270	4	US-09-581-909-3	Sequence 3, Appli
20	113	3.2	235	4	US-08-747-562-28	Sequence 28, Appl
21	112.5	3.2	745	4	US-09-107-532A-5687	Sequence 5687, Ap
22	112	3.1	519	3	US-08-997-445D-2	Sequence 2, Appli
23	109	3.1	644	1	US-08-021-608D-2	Sequence 2, Appli
24	109	3.1	644	1	US-08-726-160-2	Sequence 2, Appli
25	109	3.1	644	5	PCT-US94-01782-2	Sequence 2, Appli
26	108.5	3.1	428	1	US-08-476-008-42	Sequence 42, Appl
27	108.5	3.1	428	1	US-08-306-063-42	Sequence 42, Appl
28	108.5	3.1	428	1	US-08-833-485-42	Sequence 42, Appl
29	108.5	3.1	428	3	US-09-137-440-42	Sequence 42, Appl
30	108.5	3.1	2504	4	US-09-328-352-5821	Sequence 5821, Ap
31	108	3.0	693	1	US-08-463-620-11	Sequence 11, Appl
32	108	3.0	693	2	US-08-224-917-11	Sequence 11, Appl
33	108	3.0	693	2	US-08-914-853-11	Sequence 11, Appl
34	108	3.0	693	5	PCT-US95-03934A-11	Sequence 11, Appl
35	108	3.0	959	4	US-09-914-259-67	Sequence 67, Appl
36	107.5	3.0	961	4	US-09-914-259-66	Sequence 66, Appl
37	107	3.0	836	1	US-08-426-627-6	Sequence 6, Appli
38	106.5	3.0	837	1	US-08-426-627-23	Sequence 23, Appl
39	106.5	3.0	1118	3	US-09-379-523-3	Sequence 3, Appli
40	106	3.0	579	4	US-09-643-597-348	Sequence 348, App
41	106	3.0	579	4	US-09-542-615A-348	Sequence 348, App
42	106	3.0	579	4	US-09-606-421B-348	Sequence 348, App
43	105	3.0	444	4	US-09-107-532A-3661	Sequence 3661, Ap
44	105	3.0	533	3	US-09-603-185-2	Sequence 2, Appli
45	105	3.0	937	1	US-08-253-155A-31	Sequence 31, Appl

ALIGNMENTS

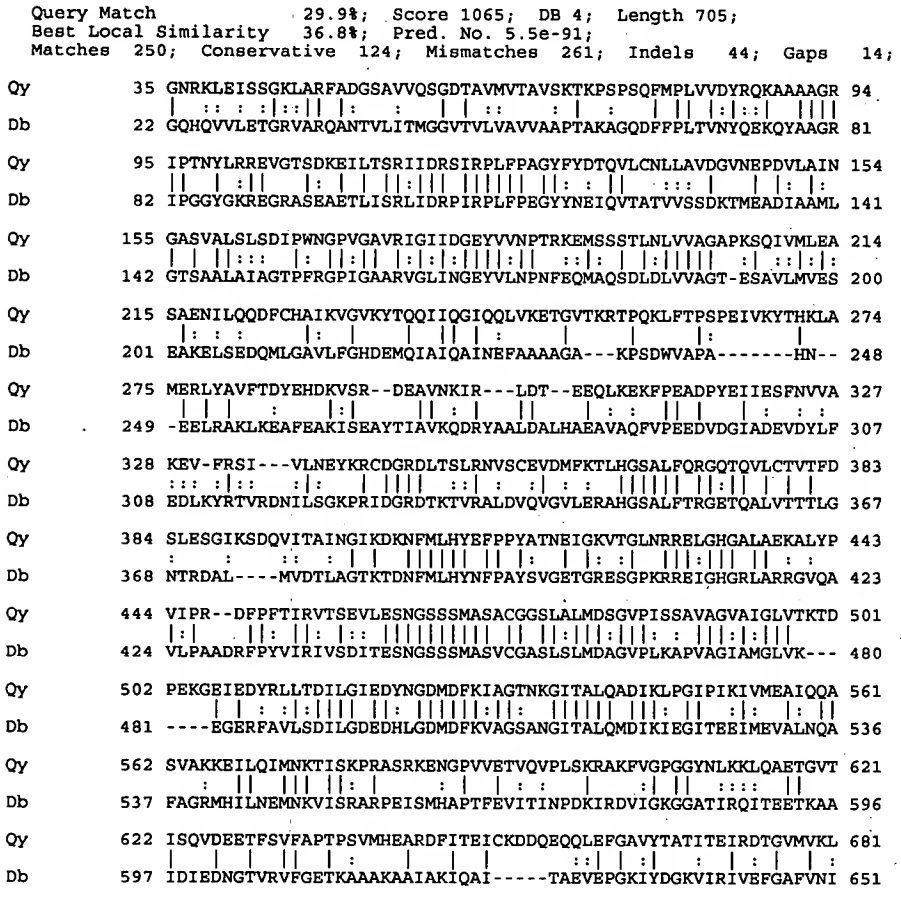
RESULT 1
US-09-134-001C-4908
; Sequence 4908, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4908
; LENGTH: 706
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4908

Query Match 32.8%; Score 1165.5; DB 4; Length 706;
Best Local Similarity 38.7%; Pred. No. 2e-100;
Matches 270; Conservative 123; Mismatches 250; Indels 55; Gaps 13;

```
Qy          24 SAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVV 83
               |  :    :   | | | :|:||: |:|: :|: ||| |: || :  :|    | || |
Db           7 SQEKKVFKTEWAGRSLTIETGQLAKQANGAVLVRYGDTVVLSTATASKEPRDGDFFPLTV 66

Qy          84 DYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVD 143
                 :| :|  |||:||  : :||    |:  ||:|:||| |||||| || :| |::  :|: |
Db          67 NYEEKMYAAGKIPGGFKKREGRPGDEATLTARLIDRPIRPLFPKGYRHDVQIMNIVLSAD 126

Qy         144 GVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAG 203
                   |:: |: |:|:|||:||||: ||:  | :| |||:||:||:  :   | |:| |||
Db         127 PDCSPEMAAMIGSSMALSVSDIPFQGPIAGVNVGYIDGKYVINPSVADKEISRLDLEVAG 186

Qy         204 APKSQIVMLEASAENILQQDFCHAIKVG-------VKYTQQIIQGIQQLVKETGVTKRTP 256
                 |  : |:|| | | : :   || |       | : |:||  || : :|
Db         187 -HKDAVNMVEAGASEITESEMLEAIFFGHEEIKRLVAFQQEIIDHIQPIKQE-------- 237

Qy         257 QKLFTP---SPEIVKYTHKLAMER--LYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKF 311
                 | |      ::|:    |  ::      | |    ||  ||| :: ::   || :
Db         238 ---FVPVERDEDLVEKVKSLTEDKGLKDTVLT---FDKQQRDENLDALK---EEVVGHFL 288

Qy         312 PEADP------YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHG 365
                 | ||       |:    | : ||  | :: :|  | ||| :  :| :  || :    ||
Db         289 DEEDPENETLVKEVYAILNDLIKEEVRRLIADEKIRPDGRKVDEIRPLESEVGLLPRAHG 348
```

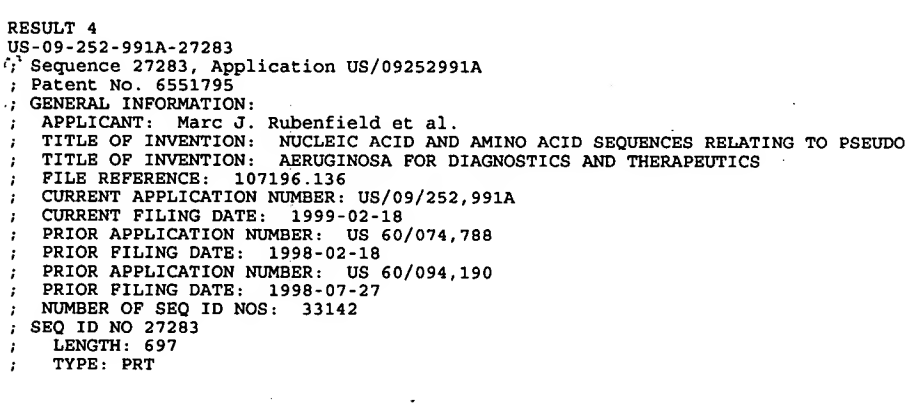
```
Qy        366 SALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTG 425
Db        349 SGLFTRGQTQALSVLTLGAL----GDYQLIDGLGPEVEKRFMHHYNFPNFSVGETGPVRA 404

Qy        426 LNRRELGHGALAEKALYPVIP--RDFPPTIRVTSEVLESNGSSSMASACGGSLALMDSGV 483
Db        405 PGRREIGHGALGERALRYIIPDTQDFPYTIRIVSEVLESNGSSSQASICGSTLALMDAGV 464

Qy        484 PISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQAD 543
Db        465 PIKAPVAGIAMGLVTRDD-------SYTILTDIQGMEDALGDMDFKVAGTKDGITAIQMD 517

Qy        544 IKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKF 603
Db        518 IKIDGLTREVIEEALEQARQGRLAIMDHMLHTIEQPREELSAYAPKVVTMSINPDKIRDV 577

Qy        604 VGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGA 663
Db        578 IGPGGKKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAKEIIEDITR-----EAEVGQ 632

Qy        664 VYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERLN 701
Db        633 VYHAKVKRIEKYGAFVELFPGKDA-LLHISQISQERIN 669


RESULT 2
US-09-107-532A-5868
; Sequence 5868, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 5868:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 750 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...750
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5868:
US-09-107-532A-5868

  Query Match             31.1%;  Score 1107;  DB 4;  Length 750;
  Best Local Similarity   36.8%;  Pred. No. 7e-95;
  Matches  250;  Conservative  144;  Mismatches  247;  Indels   38;  Gaps   12;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGR 94
Db         54 GGRPLQIEVGQLAKQANGAVLVRYGDTVVLSAAVASKEAKDTDFFPLTINYEEKMYAAGK 113

Qy         95 IPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAIN 154
Db        114 IPGGFIKRESRPSTEATLTARLIDRPIRPMFAEGFRNEVQVTNIVMSVKTDCSPAMAAML 173

Qy        155 GASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEA 214
Db        174 GSSLALSISDIPFDGPIAGVEVGRVNGEYVLNPTVEQAEQTDIELTVAGT-KQAINMVES 232

Qy        215 SAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLA 274
Db        233 GAKEVSEEDMLGALLFGFDAIKELVAFQEEIVQAVG--KEKMEVTLLQVDEVLK---KEI 287

Qy        275 MERLY-----AVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFP--EADPY---EIIESFN 324
Db        288 FDASYATMKAAVMTE---EKLAREDNIEQVKIDIREAYAEKFAGHEDEDHLLKEVKQITE 344

Qy        325 VVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALPQRGQTQVLCTVTFDS 384
Db        345 DLEKDVVRELITIDKIRPDGRKLDEIRPLSSEVSLLPRVHGSGLFTRGQTQALSACTL-- 402

Qy        385 LESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPV 444
Db        403 --APLGEHQIIDGLGVEVSKRFIHHYNFPQFSVGSTGRAGSPGRREIGHGALGERALAQV 460

Qy        445 IP--RDFPPTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDP 502
Db        461 IPSEEEFPYTIRLVAEVLESNGSSSQASICAGTLALMDAGVPIKAPVAGIAMGLVS---- 516

Qy        503 EKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQAS 562
Db        517 ---DGENYTILTDIQGLEDHLGDMDFKVAGTKDGITALQMDIKIQGITEQILTEALTQAK 573

Qy        563 VAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTI 622
Db        574 QARMEILEELTSTIAAPREELSQYAPKIEMIQIEPAKIKDVIGKGGDTINGIIDETGVKI 633

Qy        623 SQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLY 682
Db        634 DIDQDGKVSIASADAEMIKKAIKIIEDLTK-----EVKVGEVYLGKVVRIEKFGAFVNLI 688

Qy        683 PNMTAVLLHNTQLDNERLN 701
Db        689 KGKDG-LVHISQLANDRVN 706


RESULT 3
US-09-328-352-8112
; Sequence 8112, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8112
;   LENGTH: 705
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-8112
```

Query Match 29.9%; Score 1065; DB 4; Length 705;
Best Local Similarity 36.8%; Pred. No. 5.5e-91;
Matches 250; Conservative 124; Mismatches 261; Indels 44; Gaps 14;



RESULT 4
US-09-252-991A-27283
; Sequence 27283, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27283
; LENGTH: 697
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27283

Query Match 27.5%; Score 976.5; DB 4; Length 697;
Best Local Similarity 35.6%; Pred. No. 1.1e-82;
Matches 228; Conservative 130; Mismatches 230; Indels 53; Gaps 15;



RESULT 5
US-09-198-452A-1073
; Sequence 1073, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1073
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...568
; OTHER INFORMATION: Xaa=unknown or other

US-09-198-452A-1073

Query Match 24.9%; Score 884; DB 4; Length 568;
Best Local Similarity 36.9%; Pred. No. 3.9e-74;
Matches 205; Conservative 112; Mismatches 198; Indels 40; Gaps 13;

```
Qy        157 SVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASA 216
              | ||::|||| :  |  |||| || ::|:|||: |::||||:||:||  :: |:|:|
Db          8 SAALAISDIPQSNIVAGVRIGCIDNQWVINPTKTELASSTLDLVLAGT-ENAILMIEGHC 66

Qy        217 ENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAME 276
              :  ::    ||: | |:   | : :|   :| | :|        |: |::    : | :
Db         67 DFFTEEQVLDAIEFGHKHIVTICKRLQLWQEEVGKSKNLSAVYPLPA-EVLTAVKECAQD 125

Qy        277 RLYAVFTDYEHDKVSRDEAVNKIRL-DTEEQLKEKFPEADPYEIIESFNVVA------KE 329
              :  :|        :|: |:    : || : ||    |  ::  |||: |       :
Db        126 KFTELFN-------IKDKKVHAATAHEIEENILEKLQREDD-DLFSSFNIKAACKTLKSD 177

Qy        330 VFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVL--CTVTFDSLES 387
               |::: :   | ||| ||::| :: |       ||| || ||:|| |  ||:  :::
Db        178 TMRALIRDREIRADGRSLTTVRPITIETSYLPRTHGSCLFTRGETQTLAVCTLGSEAMA- 236

Qy        388 GIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP- 446
                  |   :||     | | | |||::  |:|::    |||:||| ||||||   :|
Db        237 -----QRYEDLNGEGLSKFYLQYFFPPFSVGEVGRIGSPGRREIGHGKLAEKALSHALPD 291

Qy        447 -RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKG 505
               ||:|||: | : |||||||||| ||| |||||:|||||| :||:|:||:     ::|
Db        292 SATFPYTIRIESNITESNGSSSMASVCGGCLALMDAGVPISSPIAGIAMGLIL---DDQG 348

Qy        506 EIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAK 565
               |    :|:|| |:|| |||||||||: ||||| | |||: ||   |: :|: ||
Db        349 AI----ILSDISGLEDHLGDMDFKIAGSGKGITAFQMDIKVEGITPAIMKKALSQAKQGC 404

Qy        566 KEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQV 625
               :|| |||: :| |:|    :  | :||:|:  :| |  :|||  ::::  |||| |
Db        405 NDILNIMNEALSAPKADLSQYAPRIETMQIKPTKIASVIGPGGKQIRQIIEETGVQIDVN 464

Qy        626 DEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNM 685
              |   |: | : | :::|:: |  :    ::| |  |   :| :   |  |:: |
Db        465 DLGVVSISASSASAINKAKEIIEGLV-----GEVEVGKTYRGRVTSVVAFGAFVEVLPGK 519

Qy        686 TAVLLHNTQLDNERL 700
               | | ::   :|:
Db        520 EG-LCHISECSRQRI 533
```

RESULT 6
US-09-198-452A-1074
; Sequence 1074, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1074
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...127
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-1074

Query Match 6.4%; Score 226; DB 4; Length 127;
Best Local Similarity 42.6%; Pred. No. 1.9e-13;
Matches 49; Conservative 23; Mismatches 43; Indels 0; Gaps 0;

```
Qy         39 LEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTN 98
              |   :||:|| |:|: :|:||:| |  :| :        |:|| |||::| :: |:
Db          7 LVFETGKIARQANGAVLVRSGETCVFASACAVDLDDKVDFLPLRVDYQEKFSSTGKTLGG 66

Qy         99 YLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
              :::||   |:|||| ||:||||:|| ||     | |||  : : ||   || |||
Db         67 FIKREGRPSEKEILVSRLIDRSLRPSFPYRLMQDVQVLSYVWSYDGQVLPDPLAI 121
```

RESULT 7
US-09-071-035-404
; Sequence 404, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 404:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1416 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-404

Query Match 3.9%; Score 140.5; DB 4; Length 1416;
Best Local Similarity 21.1%; Pred. No. 0.0014;
Matches 135; Conservative 79; Mismatches 236; Indels 189; Gaps 31;

```
Qy        158 VALSLSDIPWNGPVGAVRIGIIDGEY-----VVNPTRKEMS----------SSTLNLVVA 202
              : |||| |  ||  |::  ::  :|     ||||   | :          ||   : |
Db        481 INLSLS-TPVIGPNKAIQ--LVSDQYIEPISVVNPLNAETAWGNYDQNGAYSSRTTVSVM 537

Qy        203 GAPKSQIVMLEASAE--NILQ----QDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTP 256
              |: :  |  ||   :  | |      ::      |:|  ||                ||
Db        538 GSKEKPIQNLEIKVKHPNYLSLRATKEIYFYYKLGTDYTV-----------------TP 579

Qy        257 QKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADP 316
                  |    ::|:|  :  |    :  :|  | : :|::   | :||
Db        580 ----TSDGSVIKFTTPITNEIQIPIGFNYVPDSLPKDKS---IPVDT------------ 619

Qy        317 YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRN---VSCEVDMFKTLHGSALFQRGQ 373
                | : :        : |    ||   | | | :|   |:   | | :|        |
Db        620 --IPITMSAEGLTPVDTTVTTNSKRGSERTLQSSKNQFLVNARNDSFDSLSVRTKIPAG- 676
```

```
Qy        374 TQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGH 433
                  | ||  :  : :|||  :       |:      |    |   | |
Db        677 ----ADVLFDIYD--VSNDQVDSIYPQYWDRG---QYFDKPMTPNSPG------------ 715

Qy        434 GALAEKALYPVIPRD-------FPF--------------------TIRVTSEVLESNGS 465
                  || |  |        | |                      |: :|    |   :
Db        716 --------YPTITFDENTNSYTFDFGKTNKRYIIEYKNANGWIDVPTLYITGTAKEPQSN 767

Qy        466 SSMASACGGSLALMDSGVPISSA--VAGVAIGLVTKTDPEKGEIED--YRLLTDILGIED 521
              ::  ||   |::: :  : | ||   |   :  ||||      |::  :|:    :  :
Db        768 NNEGSA---SVSVQNEALDILSATQAANPTLKNVTKTTVTTKNIDNKTHRVKNPTIEL-- 822

Qy        522 YNGDMDFKIAGTNKGITALQAD---IKLPGIP--IKIVMEAIQQASVAKKEILQIMNKTI 576
                  | || |  | |   | : |:|     : :    | |  |:      | ||
Db        823 -----------TPKGTTNAQIDLNSITVKGVPEDAYSLEKTTNGAKVIFKDYTLTENITI 871

Qy        577 SKPRASRKENGPVVETV-------QVPLSKRAKFVGPGGYNLKKLQAETGVTIS------ 623
                  |          ||         |:  ||:      |       :  ||  | ::
Db        872 EYNTVSANAGQIYTETTIDSETLNQMSASKKKVTTAPITLKFSEGDAEGIVYLATATFYT 931

Qy        624 -QVDEETFSV----FAPTPSVMHEARDFITEICKDDQEQQLEFGAV----YTATITEIRD 674
                |::|  ::     |    :| | | :| |     |:: |  | |:    ||  :| :
Db        932 HNVEDENQAIAKVSFELIDNVTHTATEFTT-----DEKGQYSFDAIMTGDYTLRVTNVPQ 986

Qy        675 ---------TGVMVKLY--PNMTAVLLHNTQLDNERLNI 702
                      ||  :||    |   : |  | :|: || :
Db        987 EYSVDEEYLTGKAIKLVKGDNQLKIPLTKT-IDHSRLQV 1024
```

RESULT 8
US-09-071-035-402
; Sequence 402, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 402:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1448 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-402

Query Match 3.9%; Score 140.5; DB 4; Length 1448;
Best Local Similarity 21.1%; Pred. No. 0.0014;
Matches 135; Conservative 79; Mismatches 236; Indels 189; Gaps 31;

```
Qy        158 VALSLSDIPWNGPVGAVRIGIIDGEY-----VVNPTRKEMS----------SSTLNLVVA 202
              : |||| | ||  |::  ::  :|     ||||   | :           ||   : |
Db        508 INLSLS-TPVIGPNKAIQ--LVSDQYIEPISVVNPLNAETAWGNYDQNGAYSSRTTVSVM 564

Qy        203 GAPKSQIVMLEASAE--NILQ----QDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTP 256
              |: :  |  ||   :  | |     ::      |:|  ||                 ||
Db        565 GSKEKPIQNLEIKVKHPNYLSLRATKEIYFYYKLGTDYTV-----------------TP 606

Qy        257 QKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADP 316
                  |    ::|:|  :  |    :  :|  | : :|::   | :||
Db        607 ----TSDGSVIKFTTPITNEIQIPIGFNYVPDSLPKDKS---IPVDT------------ 646

Qy        317 YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRN---VSCEVDMFKTLHGSALFQRGQ 373
              |  : :          : |    ||    | | | :|   |:   | | :|        |
Db        647 --IPITMSAEGLTPVDTTVTTNSKRGSERTLQSSKNQFLVNARNDSFDSLSVRTKIPAG- 703

Qy        374 TQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGH 433
                  | ||  :  : :|||  :       |:      |    |   | |
Db        704 ----ADVLFDIYD--VSNDQVDSIYPQYWDRG---QYFDKPMTPNSPG------------ 742

Qy        434 GALAEKALYPVIPRD-------FPF--------------------TIRVTSEVLESNGS 465
                  || |  |        | |                      |: :|    |   :
Db        743 --------YPTITFDENTNSYTFDFGKTNKRYIIEYKNANGWIDVPTLYITGTAKEPQSN 794

Qy        466 SSMASACGGSLALMDSGVPISSA--VAGVAIGLVTKTDPEKGEIED--YRLLTDILGIED 521
              ::  ||   |::: :  : | ||   |   :  ||||      |::  :|:    :  :
Db        795 NNEGSA---SVSVQNEALDILSATQAANPTLKNVTKTTVTTKNIDNKTHRVKNPTIEL-- 849

Qy        522 YNGDMDFKIAGTNKGITALQAD---IKLPGIP--IKIVMEAIQQASVAKKEILQIMNKTI 576
                  | || |  | |   | : |:|     : :    | |  |:      | ||
Db        850 -----------TPKGTTNAQIDLNSITVKGVPEDAYSLEKTTNGAKVIFKDYTLTENITI 898

Qy        577 SKPRASRKENGPVVETV-------QVPLSKRAKFVGPGGYNLKKLQAETGVTIS------ 623
                  |          ||         |:  ||:      |       :  ||  | ::
Db        899 EYNTVSANAGQIYTETTIDSETLNQMSASKKKVTTAPITLKFSEGDAEGIVYLATATFYT 958

Qy        624 -QVDEETFSV----FAPTPSVMHEARDFITEICKDDQEQQLEFGAV----YTATITEIRD 674
                |::|  ::     |    :| | | :| |     |:: |  | |:    ||  :| :
Db        959 HNVEDENQAIAKVSFELIDNVTHTATEFTT-----DEKGQYSFDAIMTGDYTLRVTNVPQ 1013

Qy        675 ---------TGVMVKLY--PNMTAVLLHNTQLDNERLNI 702
                      ||  :||    |   : |  | :|: || :
Db       1014 EYSVDEEYLTGKAIKLVKGDNQLKIPLTKT-IDHSRLQV 1051
```

RESULT 9
US-09-252-991A-23776
; Sequence 23776, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23776
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23776

Query Match 3.9%; Score 140; DB 4; Length 252;
Best Local Similarity 22.1%; Pred. No. 7.7e-05;
Matches 62; Conservative 48; Mismatches 109; Indels 62; Gaps 11;

```
Qy        331 FRSIVLNEYK-----RCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSL 385
              :| | |::|:     |  ||    || :       |   || | : | |:|:|||   |
Db          1 WRPIPLSQYRISTMNRPSGRAADQLRPIRITRHYTKHAEGSVLVEFGDTKVICTV---SA 57

Qy        386 ESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKAL--YP 443
              |||:      : |       |  | :| |      |:||   |    : |
Db         58 ESGVP-----RFLKGQGQGWLTAEYGMLPRSTGE------RNQREASRGKQGGRTLEIQR 106

Qy        444 VIPRDF----------PFTIRVTSEVLESNGSSSMASACGGSLALMDS----------G 482
              :| |              |: :  :|::::| :  ||  | ::||:|:          |
Db        107 LIGRSLRAALDLSKLGENTLYIDCDVIQADGGTRTASITGATVALIDALAVLKKRGALKG 166

Qy        483 VPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQA 542
              |:   || |::|:                 : |:  :||   : |  :  |: |   ::
Db        167 NPLKQMVAAVSVGIYQGVP----------VLDLDYLEDSAAETDLNVVMTDAG-GFIEV 214

Qy        543 DIKLPGIP-----IKIVMEAIQQASVAKKEILQIMNKTISK 578
                  | |      :  ::|  ||     :|: ::    :::
Db        215 QGTAEGAPFRPAELNAMLELAQQ---GMQELFELQRAALAE 252


RESULT 10
US-09-252-991A-27186
; Sequence 27186, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 27186
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27186

  Query Match             3.9%;  Score 137;  DB 4;  Length 191;
  Best Local Similarity   35.5%;  Pred. No. 9.1e-05;
  Matches   33;  Conservative   18;  Mismatches   38;  Indels    4;  Gaps    2;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
              |   : : :|::|| | |: :|    | :|:|| |     |:  : | || | |::|  |||
Db         94 GQSTVTLETGRIARQATGAVLVTMDDVSVLVTVVGAKSPAEGRDFFPLSVHYQEKTYAAG 153

Qy         94 RIPTNYLRREVGTSDK---EILTSRIIDRSIRP 123
              |||  : :||   |:|   ::   |  | | |
Db        154 RIPGGFFKREGRPSEKGNPDLAPDRPSDPSAVP 186


RESULT 11
US-09-328-352-5628
; Sequence 5628, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5628
;   LENGTH: 244
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
```

```
US-09-328-352-5628

  Query Match             3.6%;  Score 128.5;  DB 4;  Length 244;
  Best Local Similarity   21.2%;  Pred. No. 0.00088;
  Matches   61;  Conservative   43;  Mismatches  109;  Indels   75;  Gaps    9;

Qy        336 LNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVI 395
              |     | | | |  || |     :  :    || | : | |:|||| : |        : |
Db          3 LGRIMRIDQRALDQLREVKITRNYTRYAEGSVLVEFGHTKVLCTASID--------NSVP 54

Qy        396 TAINGIKDKNFMLHYEFPPYATNE-------IGKVTGLN---RRELGHGALAEKALYPVI 445
              : |          |   | :|:         || ||     :| :|    |   |  :
Db         55 RFLKGQGQGWVTAEYGMLPRSTHSRCDREAARGKQTGRTQEIQRLIGRSLRAMVDLKKLG 114

Qy        446 PRDFPFTIRVTSEVLESNGSSSMASACGGSLALMDS----------GVPISSAVAGVAI 494
                  || :  :|::::| :  ||  | ::||:|:              |:   || :::
Db        115 EN----TITIDCDVIQADGGTRTASITGAAVALVDAMNVLLAQKKIKQDPLKGLVAAISV 170

Qy        495 GLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIV 554
              |:           :|  :| |:   || |  |  :  |  |
Db        171 GM----------YQD-EVLLDLCYEEDSNCQTDLNVVMTQAG----------------- 201

Qy        555 MEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAK 602
                      | ::|      ||  :| ::  ::|  :  :::  |
Db        202 -----------EFIEIQGTAEDKP-FTRAQSNAMLEMAEKGIAELIK 236


RESULT 12
US-09-134-001C-3979
; Sequence 3979, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3979
;   LENGTH: 871
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3979

  Query Match             3.6%;  Score 127.5;  DB 4;  Length 871;
  Best Local Similarity   20.0%;  Pred. No. 0.0099;
  Matches  145;  Conservative  116;  Mismatches  276;  Indels  189;  Gaps   33;

Qy         10 DRALTQLQVRALWSSAGSR--------AVAVDLGNRKLEISSGKLARFADG-SAVVQSGD 60
              |: : :| : || : | |         ||       :::: | |::  | |    :| :|
Db        235 DKTIFELDLSALVAGAKYRGEFEERLKAVL-----KEVKESEGRIILFIDEIHMLVGAGK 289

Qy         61 TAVMVTAVSKTKPSPSQ------FMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTS 114
              |   : | :  ||  ::          : :||:       :   :  ::|| |: :: :
Db        290 TDGAMDAGNMLKPMLARGELHCIGATTLNEYREYIEKDSALERRF--QKVGVSEPDVENT 347

Qy        115 RIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSD-------IPW 167
                | | ::  :   :                ||    |   :      |  |||       :|
Db        348 ISILRGLKERYEVYH-------------GVRIQDRALV----AAAELSDRYITDRFLPD 389

Qy        168 NGPVGAVRIGIID-------GEYVVNPT----------RKEMSSSTLNLVVAGAPKSQIV 210
                    | ::|         |   |||           : |:  | |       | | ::
Db        390 KA------IDLVDQACATIRTEMGSNPTELDQVNRRVMQLEIEESALKNESDNASKHRLE 443

Qy        211 MLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYT 270
                |:    |  :::   ::|  |:  :: |  :|:    |   :::  :   |
Db        444 ELQEELSN--EKEKQSSLKSRVEQEKEKIAKVQEKRAELDSSRQALEDAQTEG------N 495
```

```
Qy        271 HKLAMERLYAVFTDYEHDKVSRDEAV-NKIRLDTEEQLKEKFPEADPYEIIESFNVVAKE 329
Db        496 LEKAAELQYGTIPQLEKELQEFEEAFQDETGEDSERMIREVVSDEEIGDIVSQWT----- 550

Qy        330 VFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGI 389
Db        551 ---GIPVSKLVETEREKLLSLSDI---------LHKRVV---GQDKAVDLV--------- 586

Qy        390 KSDQVITAINGIKDKNFMLHYEFPPYATNEIGKV-----TGLNRRELGHGALAEKALYPV 444
Db        587 -SDAVVRARAGIKDPN------------RPIGSFLFLGPTGVGKTELA------KSLASS 627

Qy        445 IPRDFPPFTIRV-TSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPE 503
Db        628 LFDSEKHMIRIDMSEYMEKHAVSRLIGAPPGYVG-HDEGGQLTEAVRRNPYSVILLDEVE 686

Qy        504 KGEIEDYRLLTDILG---IEDYNG-DMDFK----IAGTNKGITALQADIKLPG----IPI 551
Db        687 KAHSDVFNVLLQILDEGRLTDSKGRSVDFKNTIIIMTSNIGSQVLLENVKDAGEISDDTE 746

Qy        552 KIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR----------- 600
Db        747 KAVMDSLH--AYFKPEILNRMDDIVLFKPLSVDDMSMIVDKILTQLNMRLLDQHISIEVT 804

Qy        601 ---AKFVGPGGYN-------LKKL---QAETGV-----------TISQVD-----EETF 630
Db        805 EEAKKWLGEEAYEPQFGARPLKRFVQRQIETPIARMMIKESLPEGTIIKVDLNDNKELDF 864

Qy        631 SVFAPT 636
Db        865 KVVKPT 870
```

RESULT 13
US-09-107-532A-5705
; Sequence 5705, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5705:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...471
; SEQUENCE DESCRIPTION: SEQ ID NO: 5705:
US-09-107-532A-5705

Query Match 3.5%; Score 124.5; DB 4; Length 471;
Best Local Similarity 21.1%; Pred. No. 0.0065;
Matches 66; Conservative 56; Mismatches 106; Indels 85; Gaps 15;

```
Qy         37 RKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIP 96
Db         32 RKLELQT-NVFKYPEGSVVIRFGDTTVICSATIENSVPP--FL--------RETGTGWVT 80

Qy         97 TNY--LRREVGTSDKE-----------ILTSRIIDRSIRPLFPAGYFYDTQVL--CNLLA 141
Db         81 AEYSMLPRATSTRNRRESSKGKLFGRTMEIQRLIGRSLRAVVDLEKLGERSIIVDCDVIQ 140

Qy        142 VDGVNEPDVLAINGASVAL-----------SLSDIPWNGPVGAVRIGII-DGEYVVNPTR 189
Db        141 ADGGTR--TASITGAFVALKLAIEKLLREKELSEDPIKEHLAAVSVGILPDGTCVTDLDY 198

Qy        190 KEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQG---IQQLV 246
Db        199 QEDSAALVDMNLVMTESGKFVEIQGTGE--------EATPDGEQLNEMLFFGKNAIEDLI 250

Qy        247 KETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTE-- 304
Db        251 KE--------QK-------------------HALLTEFAQND-ERIEETKTIIIATRNP 281

Qy        305 ---EQLKEKFPEA 314
Db        282 GKAEEFRNMFKEA 294
```

RESULT 14
US-08-999-774A-13
; Sequence 13, Application US/08999774A
; Patent No. 6274312
; GENERAL INFORMATION:
; APPLICANT: Gish, Kurt C.
; APPLICANT: Seghezzi, Wolfgang
; APPLICANT: Shanahan, Frances
; APPLICANT: Lees, Emma M.
; APPLICANT: McClanahan, Terrill K.
; TITLE OF INVENTION: Intracellular Regulatory Molecules;
; TITLE OF INVENTION: Related Reagents
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,774A
; FILING DATE: 10-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,818
; FILING DATE: 11-DEC-1996
; ATTORNEY/AGENT INFORMATION:

```
;      NAME:  Ching, Edwin P.
;      REGISTRATION NUMBER:  34,090
;      REFERENCE/DOCKET NUMBER:  DX0646
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650)852-9196
;      TELEFAX:  (650)496-1200
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1306 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-999-774A-13

  Query Match             3.5%;  Score 124.5;  DB 3;  Length 1306;
  Best Local Similarity   19.7%;  Pred. No. 0.038;
  Matches  149;  Conservative  112;  Mismatches  239;  Indels  257;  Gaps   38;

Qy          73 PSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKE-ILTSRIIDR-------SIRPL 124
               |  :| :||::             ||          |:|  |  ::|||  .      :||
Db         530 PLFNQILPLLMS-----------PT--------LEDQERHLLVKVIDRILYKLDDLVRP- 569

Qy         125 FPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDG--E 182
                   |  :   |:   | :|   |     :  |  :  :|:       : :     ||   |
Db         570 ----YVHKILVVIEPLLID---EDYYARVEGLEIISNLAKAAGLATMISTMRPDIDNMDE 622

Qy         183 YVVNPTRKEMSSSTLNLVVA---GAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQ-- 237
               || | | :  :      |||   | | | :  |:|  ::   :    |   |:|  ||
Db         623 YVRNTTARAFA------VVASALGIP-SLLPFLKAVCKS---KKSWQARHTGIKIVQQIA 672

Qy         238 ------IIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVS 291
                   |:  :: ||                    ||:                  ||  |
Db         673 ILMGCAILPHLRSLV------------------EII------------------EHGLVD 696

Qy         292 RDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAK---------------------- 328
                 :| |       | |    | ||  ||||: | |
Db         697 EQQKVRTISALAIAALAE---AATPYG-IESFDSVLKPLWKGIRQHRGKGLAAFLKAIGY 752

Qy         329 ----------------------------EVFRSIVLNEYKRCDGRDLTSLRNVSCEV-- 357
                                          |  : |||   |:| | |      :  |:
Db         753 LIPLMDAEYANYYTREVMLILIREFQSPDEEMKKIVLKVVKQCCGTDGVEANYIKTEILP 812

Qy         358 DMFKTL--HGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAI-NGIKDKNFMLHYEFPP 414
                 ||    |  |:  |   : |   | :   : : : ::|: | : :|:       ::
Db         813 PFFKHFWQHRMAILDRRNYRQLVDTTVELASNKVGAAEIISRIVDDLKDE----AEQYRK 868

Qy         415 YATNEIGKVTG-LNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASA-- 471
                  | |:  | |   ::  |  | |: :  ::   : |  : | : :  ||  :: :|
Db         869 MVMETIEKIMGNLGAADIDH-KLEEQLIDGIL---YAFQEQTTEDSVMLNGFGTVVNALG 924

Qy         472 ----------CGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDI-LGIE 520
                       ||  |  :::    |: |   |  |:::|       :: :|:  : : :
Db         925 KRVKPYLPQICGTVLWRLNNK---SAKVRQQAADLISRTAVVMKTCQEEKLMGHLGVVLY 981

Qy         521 DYNGDMDFKIAGTNKGITALQADIKLPGI-----PIKIVM-----------EAIQQ---- 560
                :| |:   :: |:  |  ||:| : : |:      |||  ::           | :|:
Db         982 EYLGEEYPEVLGSILG--ALKAIVNVIGMHKMTPPIKDLLPRLTPILKNRHEKVQENCID 1039

Qy         561 ----------ASVAKKEILQIMNKTISKPRASRKE-NGPVVETVQVPLSKRAKFVGPGGY 609
                         |: :| ::|  : :   :| :|      | |          || :||
Db        1040 LVGRIADRGAEYVSAREWMRICFELLELLKAHKKAIRRATVNT----FGYIAKAIGPHDV 1095

Qy         610 ------NLKKLQAE----TGVTISQVDEETFSVFAPTPSVMHEAR-------------- 644
                    |||  : :    | | |: |  || | |    |::|:| | |
Db        1096 LATLLNNLKVQERQNRVCTTVAIAIV-AETCSPFTVLPALMNEYRVPELNVQNGVLKSLS 1154

Qy         645 ---DFITEICKDDQEQQLEFGAVYTATITEIRDTGVM 678
                  ::| |: ||          |   :| : :  :|
Db        1155 FLFEYIGEMGKD-----------YIYAVTPLLEDALM 1180


RESULT 15
```

```
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
;   APPLICANT:  Scott, John D.,
;   APPLICANT:  Nauert, Brian J.,
;   APPLICANT:  Klauck, Theresa M.
;   TITLE OF INVENTION:  Protein Binding Domains of Gravin
;   NUMBER OF SEQUENCES:  24
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;     STREET:  6300 Sears Tower/233 South Wacker Drive
;     CITY:  Chicago
;     STATE:  Illinois
;     COUNTRY:  United States of America
;     ZIP:  60606-6402
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/769,309A
;     FILING DATE:
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  No. 5741890and, Greta E.
;     REGISTRATION NUMBER:  35,302
;     REFERENCE/DOCKET NUMBER:  27866/33451
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  312-474-6300
;     TELEFAX:  312-474-0448
;     TELEX:  25-3856
; INFORMATION FOR SEQ ID NO:  5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1780 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-769-309A-5

  Query Match             3.3%;  Score 116;  DB 1;  Length 1780;
  Best Local Similarity   19.6%;  Pred. No. 0.41;
  Matches  134;  Conservative   99;  Mismatches  269;  Indels  182;  Gaps   35;

Qy          40 EISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNY 99
               |:|  ::   |   | |  |  |  :     : |||     | :                |
Db         882 ELSESQVHMMA---AAVADGTRAATII----EERSPSWISASVTE----------PLEQ 923

Qy         100 LRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDG---VNEPDVL--AIN 154
                :  |     :|:|   :|      |          |: |       |   |:| ::   |:
Db         924 VEAEAALLTEEVLEREVIAEEEPPTV-------TEPLPENREARGDTVVSEAELTPEAVT 976

Qy         155 GASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEA 214
                | |          ||:|:                |     :  | :|       | :  |
Db         977 AAETA---------GPLGS----------EEGTEASAAEETTEMV------SAVSQLTD 1010

Qy         215 SAENILQQDFCHAIKVGV-------KYTQQIIQGIQQLVKETGVTKRT--PQKLFTPSPE 265
               | :   :      :: ||       : ||:::| : : |||      |  |: : |
Db        1011 SPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDVLQP--- 1067

Qy         266 IVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEII----- 320
               : :   :  |:  |      | | |         :::| :|   | | :    :::
Db        1068 VQRAEAERPEEQAEASGLKKETDVV--------LKVDAQEAKTEPFTQG---KVVGQTTP 1116

Qy         321 ESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTV 380
               |||   | :|  ||  :|                :|: :   || |      : | |
Db        1117 ESFE-KAPQVTESIESSELV------------TTCQAE---TLAG----VKSQEMV---- 1152

Qy         381 TFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGH-----GA 435
                   :|  | | | |  :   | :  :  :|    | :  ::  ::      |     |
Db        1153 ----MEQAIPPDSVETPTDSETDGSTPV-ADFDAPGTTQKDEIVEIHEENEVHLVPVRGT 1207
```

```
Qy        436 LAE----KALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAG 491
              ||.    :   | | | |.        |:   | |      :|   |  | : :
Db       1208 EAEAVPAQKERPPAPSSFVFQ-------EETKEQSKMED----TLEHTDKEVSVET---- 1252

Qy        492 VAIGLVTKTDPEKGEIEDY--RLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGI 549
                : :::||: : | : |       |:   |   | :|  |  : : :|    :: | |
Db       1253 --VSILSKTEGTQ-EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVT----EVALKG- 1304

Qy        550 PIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGY 609
                    |  ::|   | : |:: :   | |         || : | : :       |
Db       1305 ------EGTEEAECKKDDALELQSHAKSPP-------SPVEREMVVQVEREKTEAEPTHV 1351

Qy        610 NLKKLQAETGVTISQ-------------VD--EETFSV-FAPTPSVMHEARDFITEICK 652
              | :||: || ||:|:                :|  :|  |:  :| | :  |      :|
Db       1352 NEEKLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQE-----EAVCT 1406

Qy        653 DDQEQQLEFGAVYTATITEIRDTG 676
              | |  |    ||   | :  |
Db       1407 KIQVQSSEASFTLTAAAEEEKVLG 1430
```

Search completed: January 8, 2004, 10:41:27
Job time : 24 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 8, 2004, 10:38:52 ; Search time 37 Seconds
(without alignments)
3839.578 Million cell updates/sec

Title: US-09-907-907A-42
Perfect score: 3557
Sequence: 1 DGPFLLPRRDRALTQLQVRA.....TAVLLHNTQLDNERLNILLP 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3557	100.0	705	11	US-09-907-907A-42	Sequence 42, Appl
2	3557	100.0	705	11	US-09-907-907A-44	Sequence 44, Appl
3	1191.5	33.5	696	9	US-09-815-242-5443	Sequence 5443, Ap
4	1191.5	33.5	698	9	US-09-815-242-12325	Sequence 12325, A
5	1158	32.6	721	9	US-09-815-242-13768	Sequence 13768, A
6	1155	32.5	734	9	US-09-815-242-10312	Sequence 10312, A
7	1134	31.9	709	9	US-09-815-242-10989	Sequence 10989, A
8	1132	31.8	705	11	US-09-907-907A-43	Sequence 43, Appl
9	1128.5	31.7	704	9	US-09-815-242-10882	Sequence 10882, A
10	1127	31.7	702	9	US-09-815-242-4975	Sequence 4975, Ap
11	1097	30.8	613	9	US-09-815-242-11755	Sequence 11755, A
12	1097	30.8	701	9	US-09-815-242-12050	Sequence 12050, A
13	1054	29.6	737	9	US-09-815-242-13699	Sequence 13699, A
14	1014.5	28.5	753	10	US-09-738-626-5665	Sequence 5665, Ap
15	999.5	28.1	738	15	US-10-156-761-10060	Sequence 10060, A
16	997.5	28.0	752	10	US-09-712-363-252	Sequence 252, App
17	831.5	23.4	688	9	US-09-815-242-11416	Sequence 11416, A
18	831.5	23.4	688	12	US-09-882-227-284	Sequence 284, App
19	829.5	23.3	688	9	US-09-815-242-11576	Sequence 11576, A
20	708	19.9	164	10	US-09-925-300-1451	Sequence 1451, Ap
21	543	15.3	188	10	US-09-764-864-1115	Sequence 1115, Ap
22	287	8.1	85	10	US-09-764-864-1533	Sequence 1533, Ap
23	175	4.9	245	11	US-09-907-907A-45	Sequence 45, Appl
24	147	4.1	451	9	US-09-815-242-10721	Sequence 10721, A
25	143	4.0	245	10	US-09-738-626-6245	Sequence 6245, Ap
26	139	3.9	921	9	US-09-800-729-199	Sequence 199, App
27	136.5	3.8	245	11	US-09-893-519A-48	Sequence 48, Appl
28	136	3.8	240	12	US-10-032-585-7234	Sequence 7234, Ap
29	136	3.8	244	15	US-10-156-761-12704	Sequence 12704, A
30	131	3.7	256	15	US-10-102-806-664	Sequence 664, App
31	131	3.7	964	15	US-10-156-761-9998	Sequence 9998, Ap
32	129.5	3.6	863	12	US-10-369-493-10297	Sequence 10297, A
33	126	3.5	691	9	US-09-815-242-4977	Sequence 4977, Ap
34	126	3.5	693	9	US-09-815-242-10481	Sequence 10481, A
35	126	3.5	816	12	US-10-369-493-18348	Sequence 18348, A
36	125.5	3.5	180	11	US-09-893-519A-47	Sequence 47, Appl
37	124.5	3.5	785	12	US-10-369-493-5133	Sequence 5133, Ap
38	124.5	3.5	859	12	US-10-369-493-25	Sequence 25, Appl
39	124.5	3.5	868	12	US-10-291-172-251	Sequence 251, App
40	123.5	3.5	868	9	US-09-800-729-106	Sequence 106, App
41	122	3.4	710	12	US-10-369-493-16635	Sequence 16635, A
42	121	3.4	550	12	US-10-369-493-1230	Sequence 1230, Ap
43	119.5	3.4	538	12	US-10-369-493-1168	Sequence 1168, Ap
44	118.5	3.3	780	12	US-10-369-493-13393	Sequence 13393, A
45	118.5	3.3	1781	9	US-09-738-877-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-907-907A-42
; Sequence 42, Application US/09907907A
; Publication No. US20030099660A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; APPLICANT: Leszcyniecka, Magdalena
; TITLE OF INVENTION: GENES DISPLAYING ENHANCED EXPRESSION DURING CELLULAR SENESCENCE A
; TITLE OF INVENTION: TERMINAL CELL DIFFERENTIATION AND USES THEREOF
; FILE REFERENCE: A34584-A-PCT-USA (070050.1664)
; CURRENT APPLICATION NUMBER: US/09/907,907A
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: US 09/243,277
; PRIOR FILING DATE: 1999-02-02
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-907A-42

Query Match 100.0%; Score 3557; DB 11; Length 705;
Best Local Similarity 100.0%; Pred. No. 3.4e-302;
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60

Qy         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120

Qy        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180
```

```
Qy        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240

Qy        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300

Qy        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360

Qy        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420

Qy        421 GKVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GKVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMD 480

Qy        481 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 540

Qy        541 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 600

Qy        601 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 660

Qy        661 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERLNILLP 705
              |||||||||||||||||||||||||||||||||||||||||||||
Db        661 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERLNILLP 705
```

RESULT 2
US-09-907-907A-44
; Sequence 44, Application US/09907907A
; Publication No. US20030099660A1
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; APPLICANT: Leszcyniecka, Magdalena
; TITLE OF INVENTION: GENES DISPLAYING ENHANCED EXPRESSION DURING CELLULAR SENESCENCE A
; TITLE OF INVENTION: TERMINAL CELL DIFFERENTIATION AND USES THEREOF
; FILE REFERENCE: A34584-A-PCT-USA (070050.1664)
; CURRENT APPLICATION NUMBER: US/09/907,907A
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: US 09/243,277
; PRIOR FILING DATE: 1999-02-02
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-907A-44

Query Match 100.0%; Score 3557; DB 11; Length 705;
Best Local Similarity 100.0%; Pred. No. 3.4e-302;
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60

Qy         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120

Qy        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180

Qy        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240

Qy        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300

Qy        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360

Qy        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420

Qy        421 GKVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GKVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMD 480

Qy        481 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 540

Qy        541 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 600

Qy        601 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 660

Qy        661 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERLNILLP 705
              |||||||||||||||||||||||||||||||||||||||||||||
Db        661 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERLNILLP 705
```

RESULT 3
US-09-815-242-5443
; Sequence 5443, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 5443
; LENGTH: 696
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5443

Query Match 33.5%; Score 1191.5; DB 9; Length 696;
Best Local Similarity 39.2%; Pred. No. 4.1e-95;
Matches 274; Conservative 117; Mismatches 249; Indels 59; Gaps 12;

```
Qy         24 SAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVV 83
              |    :   :    | | | :|:||: |:|: :|: ||| |: || :  :|     | || |
Db          2 SQEKKVFKTEWAGRSLTIETGQLAKQANGAVLVRYGDTVVLSTATASKEPRDGDFFPLTV 61

Qy         84 DYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVD 143
              :| :|  |||:||  : :||     |   ||:|:||| |||||| || :| |::  :|: |
Db         62 NYEEKMYAAGKIPGGFKKREGRPGDDATLTARLIDRPIRPLFPKGYKHDVQIMNMVLSAD 121

Qy        144 GVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAG 203
                | : |: |:|:|||:||||: ||:  | :| |||:|::||| :|   | |:| |||
Db        122 PDCSPQMAAMIGSSMALSVSDIPFQGPIAGVNVGYIDGKYIINPTVEEKEVSRLDLEVAG 181

Qy        204 APKSQIVMLEASAENILQQDFCHAIKVG-------VKYTQQIIQGIQQLVKETGVTKRTP 256
                | : |:|| |  | :|:   ||  |        | : |||:  || : :|
Db        182 -HKDAVNMVEAGASEITEQEMLEAIFFGHEEIQRLVDFQQQIVDHIQPVKQE-------- 232

Qy        257 QKLFTPS-------PEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKE 309
                | |:         |   | :  ::      |     ||  ||| :: ::   || : |
Db        233 ---FIPAERDEALVERIKSLTEEKGLKETVLTF-----DKQQRDENLDNLK---EEIVNE 281

Qy        310 KFPEADP------YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTL 363
                | ||        |:    | : ||  | :: :|  | |||    :| :  || :
Db        282 FIDEEDPENELLIKEVYAILNELVKEEVRRLIADEKIRPDGRKPDEIRPLDSEVGILPRT 341

Qy        364 HGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKV 423
                ||| || ||||| |  :|  :|       :|  :   ::| || || || ::  | | |
Db        342 HGSGLFTRGQTQALSVLTLGAL----GDYQLIDGLGPEEEKRFMHHYNFPNFSVGETGPV 397

Qy        424 TGLNRRELGHGALAEKALYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDS 481
                |||:|||| |:||  :||   |||:|||: |||||||||| || || :|||||:
Db        398 RAPGRREIGHGALGERALKYIIPDTADFPYTIRIVSEVLESNGSSSQASICGSTLALMDA 457

Qy        482 GVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQ 541
                |||| : |||:|:|||: |        | :|||| |:||  |||||:||| :||||:|
Db        458 GVPIKAPVAGIAMGLVTRED-------SYTILTDIQGMEDALGDMDFKVAGTKEGITAIQ 510

Qy        542 ADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRA 601
                |||: |:  :|: ||::||   : ||:  | :|| :||        | | |: :  |
Db        511 MDIKIDGLTREIIEEALEQARRGRLEIMNHMLQTIDQPRTELSAYAPKVVTMTIKPDKIR 570

Qy        602 KFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEF 661
              :|||| : ::  |||| :   : |  : |   :::: ||: | || :     : |
Db        571 DVIGPGGKKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAREIIEEITR-----EAEV 625

Qy        662 GAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
              |  | ||:  |   |  | |:|   | ||| :|:   |:
Db        626 GQTYQATVKRIEKYGAFVGLFPGKDA-LLHISQISKNRI 663
```

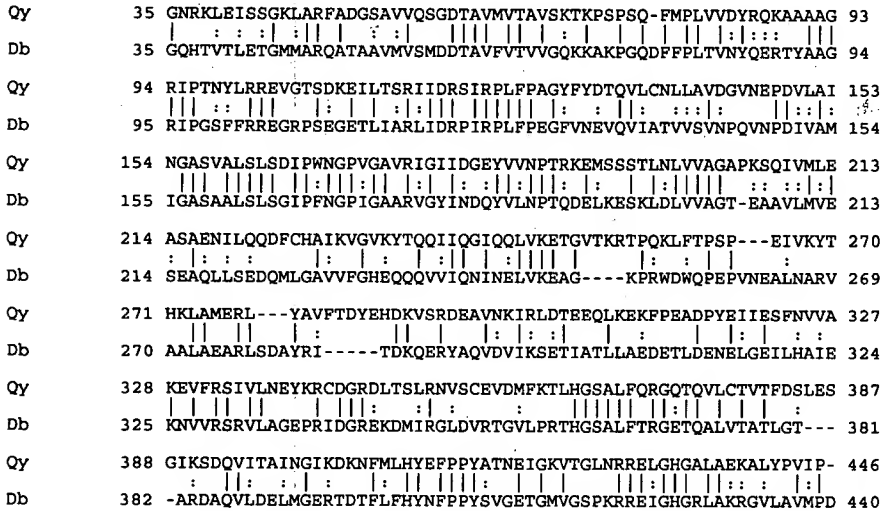
RESULT 4
US-09-815-242-12325
; Sequence 12325, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12325
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12325

Query Match 33.5%; Score 1191.5; DB 9; Length 698;
Best Local Similarity 39.2%; Pred. No. 4.1e-95;
Matches 274; Conservative 117; Mismatches 249; Indels 59; Gaps 12;

```
Qy         24 SAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVV 83
              |    :   :    | | | :|:||: |:|: :|: ||| |: || :  :|     | || |
Db          2 SQEKKVFKTEWAGRSLTIETGQLAKQANGAVLVRYGDTVVLSTATASKEPRDGDFFPLTV 61

Qy         84 DYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVD 143
              :| :|  |||:||  : :||     |   ||:|:||| |||||| || :| |::  :|: |
Db         62 NYEEKMYAAGKIPGGFKKREGRPGDDATLTARLIDRPIRPLFPKGYKHDVQIMNMVLSAD 121

Qy        144 GVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAG 203
                | : |: |:|:|||:||||: ||:  | :| |||:|::||| :|   | |:| |||
Db        122 PDCSPQMAAMIGSSMALSVSDIPFQGPIAGVNVGYIDGKYIINPTVEEKEVSRLDLEVAG 181

Qy        204 APKSQIVMLEASAENILQQDFCHAIKVG-------VKYTQQIIQGIQQLVKETGVTKRTP 256
                | : |:|| |  | :|:   ||  |        | : |||:  || : :|
Db        182 -HKDAVNMVEAGASEITEQEMLEAIFFGHEEIQRLVDFQQQIVDHIQPVKQE-------- 232

Qy        257 QKLFTPS-------PEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKE 309
                | |:         |   | :  ::      |     ||  ||| :: ::   || : |
Db        233 ---FIPAERDEALVERIKSLTEEKGLKETVLTF-----DKQQRDENLDNLK---EEIVNE 281

Qy        310 KFPEADP------YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTL 363
                | ||        |:    | : ||  | :: :|  | |||    :| :  || :
Db        282 FIDEEDPENELLIKEVYAILNELVKEEVRRLIADEKIRPDGRKPDEIRPLDSEVGILPRT 341

Qy        364 HGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKV 423
                ||| || ||||| |  :|  :|       :|  :   ::| || || || ::  | | |
Db        342 HGSGLFTRGQTQALSVLTLGAL----GDYQLIDGLGPEEEKRFMHHYNFPNFSVGETGPV 397

Qy        424 TGLNRRELGHGALAEKALYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDS 481
                |||:|||| |:||  :||   |||:|||: |||||||||| || || :|||||:
Db        398 RAPGRREIGHGALGERALKYIIPDTADFPYTIRIVSEVLESNGSSSQASICGSTLALMDA 457

Qy        482 GVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQ 541
                |||| : |||:|:|||: |        | :|||| |:||  |||||:||| :||||:|
Db        458 GVPIKAPVAGIAMGLVTRED-------SYTILTDIQGMEDALGDMDFKVAGTKEGITAIQ 510

Qy        542 ADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRA 601
                |||: |:  :|: ||::||   : ||:  | :|| :||        | | |: :  |
Db        511 MDIKIDGLTREIIEEALEQARRGRLEIMNHMLQTIDQPRTELSAYAPKVVTMTIKPDKIR 570

Qy        602 KFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEF 661
              :|||| : ::  |||| :   : |  : |   :::: ||: | || :     : |
Db        571 DVIGPGGKKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAREIIEEITR-----EAEV 625
```

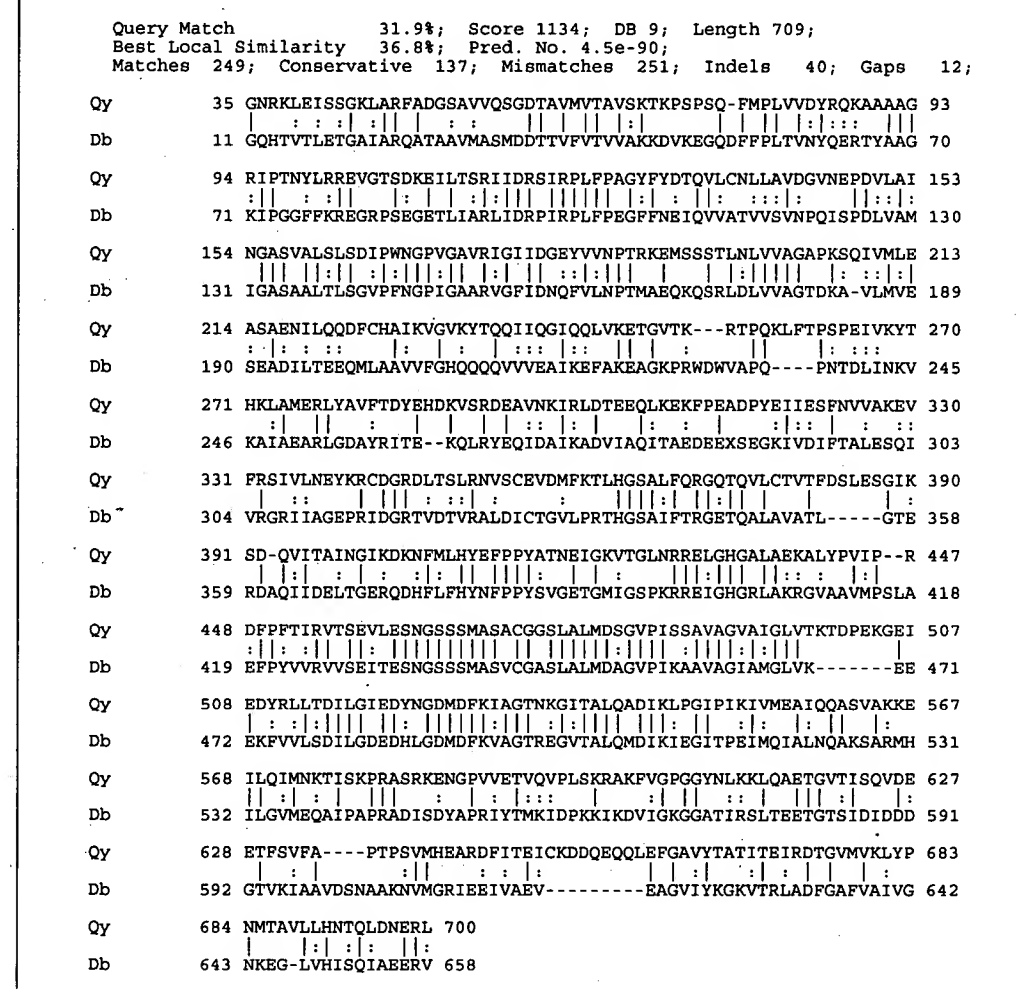
```
Qy        662 GAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
Db        626 GQTYQATVKRIEKYGAFVGLFPGKDA-LLHISQISKNRI 663


RESULT 5
US-09-815-242-13768
; Sequence 13768, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13768
;   LENGTH: 721
;   TYPE: PRT
;   ORGANISM: Salmonella typhi
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(721)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13768

  Query Match             32.6%;  Score 1158;  DB 9;  Length 721;
  Best Local Similarity   39.9%;  Pred. No. 3.7e-92;
  Matches  270;  Conservative  117;  Mismatches  250;  Indels   40;  Gaps   11;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
Db         22 GQHTXTLETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQDFFPLTVNYQERTYAAG 81

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
Db         82 RIPGSFXRREGRPSEGETLIARLIDRPVRPLFPEGFVNEVQVIATVVSVNPQVNPDIVAM 141

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
Db        142 IGASAALSLSGIPFNGPIGAARVGYINDQYVLNPTQDELKESKLDLVVAGT-EAAVLMVE 200

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVK----- 268
Db        201 SEAELLSEDTMLGAVVFGHEQQQVVIQAINDLVKEAGKPRWDWQ------PEAVNDALNA 254

Qy        269 YTHKLAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNV 325
Db        255 RVAALAESRLSDAYRI-----TDKQERYAQVDVIKSETIEQLIAEDETLDANELGEILHA 309
```

```
Qy        326 VAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSL 385
Db        310 IEKNVVRSRVLAGEPRIDGREKDMIRGLDVRTGVLPRTHGSALFTRGETQALVTATLGT- 368

Qy        386 ESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVI 445
Db        369 ---ARDAQVLDELMGERTDSFLFHYNFPPYSVGETGMVGSPKRREIGHGRLAKRGVLAVM 425

Qy        446 P--RDFPPTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPE 503
Db        426 PDMDKFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVKEGD-- 483

Qy        504 KGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASV 563
Db        484 -----NYVVLSDILGDEDHLGDMDFKVAGSRDGISALQMDIKIEGITKEIMQVALNQAKG 538

Qy        564 AKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTIS 623
Db        539 ARLHILGVMEQAINAPRGDISEFAPRIHTIKISTDKIKDVIGKGGSVIRALTEETGTTIE 598

Qy        624 QVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYP 683
Db        599 IEDDGTVKIAATDGEKAKYAIRRIEEI-----TAEIEVGRIYNSKVTRIVDFGAFVAIGG 653

Qy        684 NMTAVLLHNTQLDNERL 700
Db        654 GKEG-LVHISQIADKRV 669


RESULT 6
US-09-815-242-10312
; Sequence 10312, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10312
;   LENGTH: 734
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-09-815-242-10312

  Query Match             32.5%;  Score 1155;  DB 9;  Length 734;
  Best Local Similarity   39.9%;  Pred. No. 6.9e-92;
  Matches  269;  Conservative  118;  Mismatches  252;  Indels   36;  Gaps   11;
```



RESULT 7
US-09-815-242-10989
; Sequence 10989, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10989
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(709)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-10989

Query Match 31.9%; Score 1134; DB 9; Length 709;
Best Local Similarity 36.8%; Pred. No. 4.5e-90;
Matches 249; Conservative 137; Mismatches 251; Indels 40; Gaps 12;



```
RESULT 8
US-09-907-907A-43
; Sequence 43, Application US/09907907A
; Publication No. US20030099660A1
; GENERAL INFORMATION:
;  APPLICANT: Fisher, Paul B.
;  APPLICANT:  Leszcyniecka, Magdalena
;  TITLE OF INVENTION: GENES DISPLAYING ENHANCED EXPRESSION DURING CELLULAR SENESCENCE A
;  TITLE OF INVENTION:  TERMINAL CELL DIFFERENTIATION AND USES THEREOF
;  FILE REFERENCE: A34584-A-PCT-USA (070050.1664)
;  CURRENT APPLICATION NUMBER: US/09/907,907A
;  CURRENT FILING DATE:  2001-07-16
;  PRIOR APPLICATION NUMBER: US 09/243,277
;  PRIOR FILING DATE: 1999-02-02
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
;   LENGTH: 705
;   TYPE: PRT
;   ORGANISM: Bacillus subtilis
US-09-907-907A-43

  Query Match             31.8%;  Score 1132;  DB 11;  Length 705;
  Best Local Similarity   37.4%;  Pred. No. 6.7e-90;
  Matches  251;  Conservative  144;  Mismatches  253;  Indels   24;  Gaps   10;

Qy         32 VDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAAA 91
Db         10 IDWAGRTLTVETGQLAKQANGAVMIRYGDTAVLSTATASKEPKPLDFFPLTVNYEERLYA 69

Qy         92 AGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVL 151
Db         70 VGKIPGGFIKREGRPSEKAVLASRLIDRPIRPLFADGFRNEVQVISIVMSVDQNCSSEMA 129

Qy        152 AINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVM 211
Db        130 AMFGSSLALSVSDIPFEGPIAGVTVGRIDDQFIINPTVDQLEKSDINLVVAGT-KDAINM 188

Qy        212 LEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTH 271
Db        189 VEAGADEVPEEIMLEAIMFGHEEIKRLIAFQEEIVAAVG-KEKSEIKLFEIDEELNEKVK 247

Qy        272 KLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPY-EIIESFNVVAKEV 330
Db        248 ALAEEDLLKAI--QVHEKHAREDAINEVKNAVVAKFEDEEHDEDTIKQVKQILSKLVKNE 305

Qy        331 FRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIK 390
Db        306 VRRLITEEKVRPDGRGVDQIRPLSSEVGLLPRTHGSGLFTRGQTQALSVCTLGAL----G 361

Qy        391 SDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP--RD 448
Db        362 DVQILDGLGVEESKRFMHHYNFPQFSVGETGPMRGPGRREIGHGALGERALEPVIPSEKD 421

Qy        449 FPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEIE 508
Db        422 FPYTVRLVSEVLESNGSTSQASICASTLAMMDAGVPIKAPVAGIAMGLV-----KSG--E 474

Qy        509 DYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKEI 568
Db        475 HYTVLTDIQGMEDALGDMDFKVAGTEKGVTALQMDIKIEGLSREILEEALQQAKKGRMEI 534

Qy        569 LQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDEE 628
Db        535 LNSMLATLSESRKELSRYAPKILTMTINPDKIRDVIGPSGKQINKIIEETGVKIDIEQDG 594

Qy        629 TFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMTAV 688
Db        595 TIFISSTDESGNQKAKKII-----EDLVREVEVGQLYLGKVKRIEKFGAFVEIFSGKDG- 648

Qy        689 LLHNTQLDNERL 700
Db        649 LVHISELALERV 660


RESULT 9
US-09-815-242-10882
; Sequence 10882, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10882
;   LENGTH: 704
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-09-815-242-10882

  Query Match             31.7%;  Score 1128.5;  DB 9;  Length 704;
  Best Local Similarity   36.7%;  Pred. No. 1.4e-89;
  Matches  258;  Conservative  147;  Mismatches  245;  Indels   53;  Gaps   15;

Qy         13 LTQLQV-RALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKT 71
Db          1 MTEKQVFKTTW-----------GGRPLEVEIGQLAKQANGAVLVRYGDTVVLSAAVASK 48

Qy         72 KPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFY 131
Db         49 EAKDVDFFPLTVNYEEKMYAVGKIPGGFIKREGRPSERATLTARLIDRPIRPMFSEGFRN 108

Qy        132 DTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKE 191
Db        109 EVQITNIVMSVEQDCTPEMAAMFGSSLALAISDIPFDGPIAGVDVGRINGEYVLNPTVEQ 168

Qy        192 MSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGV 251
Db        169 AEQTDIELTVAGT-KEAINMVESGAKEVSEEDMLGALLFGFDAIKELVAFQEEIVAAVGK 227

Qy        252 TKRTPQKLFTPSPEIVKYTHKLAMERLY-----AVFTDYEHDKVSRDEAVNKIRLDTEEQ 306
Db        228 PKMDVDLL-----QVDADLKKEIFDAYYNTMKTAVMTE---EKLAREVEIDKVKDTVKEV 279

Qy        307 LKEKFPE-ADPYEIIESFNVVA----KEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFK 361
Db        280 YAEKFAEHEEEAQLLKEVKQIAEDLEKDVVRELITIDKIRPDGRKLDEIRHLSSEVSILP 339

Qy        362 TLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKD-KNFMLHYEFPPYATNEI 420
```

```
Db        340 RVHGSGLFTRGQTQALSVCTL----APLGEHQIIDGL-GVQDSKRFIHHYNFPQFSVGST 394

Qy        421 GKVTGLNRRELGHGALAEKALYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLAL 478
              |:     |||:|||||  |:|   :||   |||:|||: :||||||||| || | |:|||
Db        395 GRAGSPGRREIGHGALGERALAQIIPSEEDFPYTIRLVAEVLESNGSSSQASICAGTLAL 454

Qy        479 MDSGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGIT 538
              ||:|||| : |||:|:|||:       : |:| :|||| |:||: ||||||:||| |||
Db        455 MDAGVPIKAPVAGIAMGLVS-------DGENYTILTDIQGLEDHLGDMDFKVAGTKDGIT 507

Qy        539 ALQADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLS 598
              ||| |||: ||  :|: ||: ||  |: |||: :  ||:||    :  | :| :|:  :
Db        508 ALQMDIKIQGITEQILTEALDQAKKARMEILEELTTTIAAPREELSQYAPKIEMIQIKPA 567

Qy        599 KRAKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQ 658
              |    :| ||  :  :  |||| |   :   |: :    :: :|   | |: |    :
Db        568 KIKDVIGKGGETINSIIDETGVKIDIDQDGNVSIASSDAEMIKKAIKIIEELTK-----E 622

Qy        659 LEFGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERLN 701
              :| | || | :  |   | | |       |:| :|| |||:|
Db        623 VEVGQVYLAKVVRIEKFGAFVNLIKGKDG-LIHISQLANERVN 664


RESULT 10
US-09-815-242-4975
; Sequence 4975, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4975
;   LENGTH: 702
;   TYPE: PRT
;   ORGANISM: Enterococcus faecalis
US-09-815-242-4975

  Query Match             31.7%;  Score 1127;  DB 9;  Length 702;
  Best Local Similarity   37.4%;  Pred. No. 1.8e-89;
  Matches  254;  Conservative  144;  Mismatches  242;  Indels   40;  Gaps   13;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGR 94
              | | ||:  |:||: |:|: :|: ||| |:  ||:  :     | || |:| :|  | |:
Db         10 GGRPLEVEIGQLAKQANGAVLVRYGDTVVLSAAVASKEAKDVDFFPLTVNYEEKMYAVGK 69

Qy         95 IPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAIN 154
              ||  :::||   |::  ||:|:||| |||:|  |:  : |:   :::|:    |:: |:

Db         70 IPGGFIKREGRPSERATLTARLIDRPIRPMFSEGFRNEVQITNIVMSVEQDCTPEMAAMF 129

Qy        155 GASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEA 214
              |:|:||::||||::||:  | :| |:||||:||| ::   : : | |||  |  | |:|:
Db        130 GSSLALAISDIPFDGPIAGVDVGRINGEYVLNPTVEQAEQTDIELTVAGT-KEAINMVES 188

Qy        215 SAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLA 274
              |: : ::|   |: |    :::   :::|   |  |      |     ::     |
Db        189 GAKEVSEEDMLGALLFGFDAIKELVAFQEEIVAAVGKPKMDVDLL-----QVDADLKKEI 243

Qy        275 MERLY-----AVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPE-ADPYEIIESFNVVA- 327
              :  |      || |:   :|::|:  ::|::   :|   ||| |  :  ::::    :|
Db        244 FDAYYNTMKTAVMTE---EKLAREVEIDKVKDTVKEVYAEKFAEHEEEAQLLKEVKQIAE 300

Qy        328 ---KEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDS 384
                 |:| | ::  :  | ||| |  :|::| || :   :||| || ||||| |   |
Db        301 DLEKDVVRELITIDKIRPDGRKLDEIRHLSSEVSILPRVHGSGLFTRGQTQALSVCTL-- 358

Qy        385 LESGIKSDQVITAINGIKD-KNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYP 443
              :  :   |:|  : |::| | |: || || ::    |:     |||:|||| |:||
Db        359 --APLGEHQIIDGL-GVQDSKRFIHHYNFPQFSVGSTGRAGSPGRREIGHGALGERALAQ 415

Qy        444 VIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTD 501
              :||   |||:|||: :|||||||||| || | |:|||||:|||| : |||:|:|||:
Db        416 IIPSEEDFPYTIRLVAEVLESNGSSSQASICAGTLALMDAGVPIKAPVAGIAMGLVS--- 472

Qy        502 PEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQA 561
                : |:| :|||| |:||: ||||||:|||  ||||||  |||: ||  :|: ||: ||
Db        473 ----DGENYTILTDIQGLEDHLGDMDFKVAGTKDGITALQMDIKIQGITEQILTEALDQA 528

Qy        562 SVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVT 621
              |: |||: :  ||: ||     :  | :| :|:  :|     :| ||  :  :  ||||
Db        529 KKARMEILEELTTTIAAPREELSQYAPKIEMIQIKPAKIKDVIGKGGETINSIIDETGVK 588

Qy        622 ISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKL 681
              |    :   |: :    :: :|   | |: |      ::| | || | :  |   | | |
Db        589 IDIDQDGNVSIASSDAEMIKKAIKIIEELTK-----EVEVGQVYLAKVVRIEKFGAFVNL 643

Qy        682 YPNMTAVLLHNTQLDNERLN 701
                    |:| :|| |||:|
Db        644 IKGKDG-LIHISQLANERVN 662


RESULT 11
US-09-815-242-11755
; Sequence 11755, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
```

```
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11755
;   LENGTH: 613
;   TYPE: PRT
;   ORGANISM: Klebsiella pneumoniae
US-09-815-242-11755

  Query Match             30.8%;  Score 1097;  DB 9;  Length 613;
  Best Local Similarity   40.4%;  Pred. No. 6.2e-87;
  Matches  248;  Conservative  110;  Mismatches  222;  Indels   34;  Gaps    9;

Qy         39 LEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAGRIPT 97
Db          1 MTLETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQDFFPLTVNYQERTYAAGKIPG 60

Qy         98 NYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGAS 157
Db         61 GFFRREGRPSEGETLIARLIDRPVRPLFPEGFVNEVQVIATVVSVNPQVNPDIVAMIGAS 120

Qy        158 VALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAE 217
Db        121 AALSLSGIPFNGPIGAARVGYINDQYVLNPTQEELKSSKLDLVVAGT-EAAVLMVESEAE 179

Qy        218 NILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVK-----YTHK 272
Db        180 LLSEDQMLGAVVFGHEQQQIVIQNINDLVKEAGKPRWDWQ------PEAVNEALNARVAA 233

Qy        273 LAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAKE 329
Db        234 LAESRLSDAYRI-----TDKQERYAQVDVIKSETIATLVAEDETLDANELGEILHAIEKN 288

Qy        330 VFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGI 389
Db        289 VVRSRVLAGEPRIDGREKDMIRGLDVRTGVLPRTHGSALFTRGETQALVTATLGTARDAQ 348

Qy        390 KSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP--R 447
Db        349 NIDELM----GERTDSFLFHYNFPQYSVGETGMVGSPKRREIGHGRLAKRGVLAVMPTIE 404

Qy        448 DFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEI 507
Db        405 EFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPVKAAVAGIAMGLVKEGD------ 458

Qy        508 EDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKE 567
Db        459 -NFVVLSDILGDEDHLGDMDFKVAGSRDGISALQMDIKIEGITKEIMQVALNQAKGARLH 517

Qy        568 ILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDE 627
Db        518 ILGVMEQAINAPRGDISEFAPRIHTIKINPDKIKDVIGKGGSVIRALTEETGTTIEIEDD 577

Qy        628 ETFSVFAPTPSVMH 641
Db        578 GTVKIAATDGDKAH 591


RESULT 12
US-09-815-242-12050
; Sequence 12050, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12050
;   LENGTH: 701
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12050

  Query Match             30.8%;  Score 1097;  DB 9;  Length 701;
  Best Local Similarity   36.5%;  Pred. No. 7.7e-87;
  Matches  248;  Conservative  141;  Mismatches  250;  Indels   40;  Gaps   14;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
Db         11 GQSTVTLETGRIARQATGAVLVTMDDVSVLVTVVGAKSPAEGRDFFPLSVHYQEKTYAAG 70

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
Db         71 RIPGGFFKREGRPSEKETLTSRLIDRPIRPLFPEGFMNEVQVVCTVVSTNKKSDPDIAAM 130

Qy        154 NGASVALSLSDIPWNGPVGAVRIGI-IDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVML 212
Db        131 IGTSAALAISGIPFAGPIGAARVGFHPEIGYILNPTYEQLQSSSLDMVVAGT-EDAVLMV 189

Qy        213 EASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTK---RTPQKLFTPSPEIVKY 269
Db        190 ESEADELTEDQMLGAVLFAHDEFQAVIRAVKELAAEAGKPAWDWKAPAE-NTVLVNAIKA 248

Qy        270 THKLAMERLYAVFTDYEHDKVS-----RDEAVNKIRLDTEEQLKEKFPEADPYEIIESFN 324
Db        249 ELGEAISQAYTI--TIKQDRYNRLGELRDQAVALFAGEEE----GKFPAS---EVKDVFG 299

Qy        325 VVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDS 384
Db        300 LLEYRTVRENIVNGKPRIDGRDTRTVRPLRIEVGVLGKTHGSALFTRGETQALVVATLGT 359

Qy        385 LESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPV 444
Db        360 ----ARDAQLLDTLEGERKDAFMLHYNFPPFSVGECGRMGSPGRREIGHGRLARRGVAAM 415

Qy        445 IPR--DFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDP 502
Db        416 LPTQDEFPYTIRVVSEITESNGSSSMASVCGASLALMDAGVPVKAPVAGIAMGLVK---- 471

Qy        503 EKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQAS 562
Db        472 ---EGEKFAVLTDILGDEDHLGDMDFKVAGTDKGVTALQMDIKINGITEEIMEIALGQAL 528

Qy        563 VAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTI 622
Db        529 EARLNILGQMNQVIAKPRAELSENAPTMLQMKIDSDKIRDVIGKGGATIRGICEETKASI 588

Qy        623 SQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLY 682
Db        589 DIEDDGSVKIYGETKEAAEAAKLRVLAI-----TAEAEIGKIYVGKVERIVDFGAFVNIL 643
```

```
Qy        683 PNMTAVLLHNTQLDNERLN 701
              |   |:| :|: ::|::
Db        644 PGKDG-LVHISQISDKRID 661


RESULT 13
US-09-815-242-13699
; Sequence 13699, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13699
;   LENGTH: 737
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-815-242-13699

  Query Match             29.6%;  Score 1054;  DB 9;  Length 737;
  Best Local Similarity   34.6%;  Pred. No. 4.9e-83;
  Matches  234;  Conservative  148;  Mismatches  257;  Indels   38;  Gaps    9;

Qy         37 RKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIP 96
              |:| : :|::|: |:|| ||: |:: |:  ||   | :   | || |:| :|  |||: |
Db         13 RELIVETGQVAKQANGSVVVRYGESTVLTAAVMSKKMATGDFFPLQVNYEEKMYAAGKFP 72

Qy         97 TNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGA 156
                :::||   |    ||:|:||| |||:|  |:  : ||:  :|:|      : |: |:
Db         73 GGFMKREGRPSTDATLTARLIDRPIRPMFAEGFRNEVQVINTVLSYDENASAPMAAMFGS 132

Qy        157 SVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASA 216
              |:|||:||||::||:  |::| :||: ::||::::   | | | |||  |  | |:|: |
Db        133 SLALSISDIPFDGPIAGVQVGYVDGQIIINPSQEQAEQSLLELTVAGT-KHAINMVESGA 191

Qy        217 ENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAME 276
              : : ::    |:  | :  :::|   :::|   | |           |::    :|  |
Db        192 KELSEEIMLEALLKGHEAVKELIAFQEEIVAAVGKEK--------AEVELLHVDAELQAE 243

Qy        277 RLYAVFTDYE-----HDKVSRDEAVNKIRLDTEEQLKEKFPEADPY-----EIIESFNVV 326
              : |  :| :      :|::|: |   ::       :||:   : :     :: |    :
Db        244 IIAAYNSDLQKAVQVEEKLAREAATQAVKDQVTAVYEEKYANHEEFDRIMRDVAEILEQM 303

Qy        327 AKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLE 386
                 | ::  :  | ||| :  :| :   ||    :||| || ||||| |  :|
Db        304 EHAEVRRLITEDKVRPDGRKVDEIRPLDAVVDFLPRVHGSGLFTRGQTQALSVLTL---- 359

Qy        387 SGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP 446
              : :   |:|  ::     | || || || |:  | |:     |||:||||| |:||  |:|
Db        360 APMGETQIIDGLDPEYKKRFMHHYNFPQYSVGETGRYGAPGRREIGHGALGERALAQVLP 419

Qy        447 --RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEK 504
                 :||: ||: :|||||||||| || | |:|||  |||| : |||:|:||::
Db        420 SLEEFPYAIRLVAEVLESNGSSSQASICAGTLALMTGGVPIKAPVAGIAMGLIS------ 473

Qy        505 GEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVA 564
               :  :| :|||| |:||: |||||||:|||  |||||| |||: ||  :|: ||: ||  |
Db        474 -DGNNYTVLTDIQGLEDHFGDMDFKVAGTRDGITALQMDIKIQGITAEILTEALAQAKKA 532

Qy        565 KKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQ 624
              : ||| ::  || : |       | ::|::: : |    :| ||  : |: |||| |
Db        533 RFEILDVIEATIPEVRPELAPTAPKIDTIKIDVDKIKIVIGKGGETIDKIIAETGVKIDI 592

Qy        625 VDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPN 684
              :|   |:::     :: |:: |  : :: :  :     || | :  |  |  | |: :
Db        593 DEEGNVSIYSSDQDAINRAKEIIAGLVREAKVDE-----VYRAKVVRIEKFGAFVNLF-D 646

Qy        685 MTAVLLHNTQLDNERLN 701
              |  |:| :::   | |
Db        647 KTDALVHISEMAWTRTN 663


RESULT 14
US-09-738-626-5665
; Sequence 5665, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5665
;   LENGTH: 753
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-5665

  Query Match             28.5%;  Score 1014.5;  DB 10;  Length 753;
  Best Local Similarity   35.8%;  Pred. No. 1.4e-79;
  Matches  255;  Conservative  125;  Mismatches  263;  Indels   69;  Gaps   19;

Qy         33 DLGNRKLEISSGKLARFADGSAVVQ-SGDTAVMVTAVSKTKPSPS-QFMPLVVDYRQKAA 90
              | | | :   :|:||| |||:       || :: | :  :|      | || ||  ::
Db         25 DFGTRTIRFETGQLARQADGAVTTYLDDDTMLLATTTASNQPREGFDFFPLTVDVEERMY 84

Qy         91 AAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDV 150
              |||||| :: |||    | : ||  |:||| :|| |  |   :  |::  ::::  :  ||
Db         85 AAGRIPGSFFRREGRPSTEAILACRLIDRPLRPTFVKGLRNEVQIVVTVMSMNPEDYYDV 144

Qy        151 LAINGASVALSLSDIPWNGPVGAVRIGII------DGEYVVNPTRKEMSSSTLNLVVAGA 204
              :|||||| |  :| :| :| || ||: ::      :|::|  ||  :   |  :||||
```

```
Db        145 VAINGASAATRISGLPVSGAVGGVRMALVVDEKHPEGQWVAFPTHAQHEQSVFEIVVAGR 204

Qy        205 PKSQ----------IVMLEASA-ENILQQ-------------DFCHAIKVGVKYTQQI 238
Db        205 LVERKRGNKTFSDVAVMMVEAGASENVVNRVKDGAPAPTEKIVSDGLEAAKPFIDILCRA 264

Qy        239 IQGIQQLVKETGVTKRTPQKLFTP-SPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVN 297
Db        265 QEGLAQRV--GNAAKEFP--LFPPYTDEVYSAVERKVSKKLASLLT--LKAKQERDDATN 318

Qy        298 KIRLDTEEQLKEKF-------PEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSL 350
Db        319 AYMEEIEAELLPKFEASYSSAAEASK-EIRAGYNAVMKAIVRRMILTDHFRIDGRGVTDI 377

Qy        351 RNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHY 410
Db        378 RDLAVEVELIPRAHGSSLFERGETQILGVTTLDML----KMEQQIDSLAPGDAKRYMHHY 433

Qy        411 EFPPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP--RDFPFTIRVTSEVLESNGSSSM 468
Db        434 NFPPPYSTGETGRVGSPKRREIGHGALAERAVLPVIPSREEFPYAIRQVSEALGSNGSTSM 493

Qy        469 ASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEIE---DYRLLTDILGIEDYNGD 525
Db        494 GSVCASTLSLYNAGVPLKAPVAGIAMGLVS------GEIDGKTEYVALTDILGAEDAFGD 547

Qy        526 MDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKE 585
Db        548 MDFKVAGTADFITALQLDTKLDGIPSKVLSDALEQARDARLTILNTMADVINGPDEMSK- 606

Qy        586 NGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARD 645
Db        607 FAPRITTVKIPVAKIGELIGPKGKNINALTEETGANISIEDDGTVFISAADGASAEAAIE 666

Qy        646 FITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDN 697
Db        667 KINALA---NPQLPKVGERFLGTVVKTTAFGAFVSLLPGRDG-LVHISKLGN 714
```

RESULT 15
US-10-156-761-10060
; Sequence 10060, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10060
; LENGTH: 738
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10060

Query Match 28.1%; Score 999.5; DB 15; Length 738;
Best Local Similarity 34.5%; Pred. No. 2.9e-78;
Matches 240; Conservative 130; Mismatches 262; Indels 63; Gaps 16;

```
Qy         35 GNRKLEISSGKLARFADGSAVVQ-SGDTAVMVTAVSKTKPSPS-QFMPLVVDYRQKAAAA 92
Db         18 GTRTIRFETGRLAKQAAGSAVAYLDDDTMVLSATTASKKPKENLDFFPLTVDVEERMYAA 77

Qy         93 GRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLA 152
Db         78 GKIPGSFFRREGRPSEDAILTCRLIDRPLRPSFKKGLRNEIQVVATIMALNPDHLYDVVA 137

Qy        153 INGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQ---- 208
Db        138 INAASASTQLAGLPFSGPIGGVRVALINGQWVAFPTHTELEDAVFDMVVAGRALEDGDVA 197

Qy        209 IVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKE---TGVTKRTP--QKLFTPS 263
Db        198 IMMVEAEATE-------------KTIQLVAGGAEAPTEEVVAAGLDAAKPFIKVLCKAQ 243

Qy        264 PEIVKYTHKLAMERLYAVFTDYEHD--------------------KVSRDEAVNKIRLD 302
Db        244 ADLAAKAAKPTGE--FPVFLDYQDDVLEALTAAVKSELSQALTIAGKQDREAELDRVKEI 301

Qy        303 TEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKT 362
Db        302 AAEKLLPQF-EGREKEISAAYRSLTKSLVRERVIKDKVRIDGRGVTDIRTLAAEVEAIPR 360

Qy        363 LHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEPPPYATNEIGK 422
Db        361 VHGSALFERGETQILGVTTLNML----RMEQQLDTLSPVTRKRYMHNYNFPPYSVGETGR 416

Qy        423 VTGLNRRELGHGALAEKALYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMD 480
Db        417 VGSPKRREIGHGALAERAIVPVLPTREEFPYAIRQVSEALGSNGSTSMGSVCASTMSLLN 476

Qy        481 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 540
Db        477 AGVPLKAPVAGIAMGLISQ--EINGETH-YVALTDILGAEDAFGDMDFKVAGTKEFVTAL 533

Qy        541 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 600
Db        534 QLDTKLDGIPASVLAAALKQARDARLHILDVMMEAIDTP-DEMSPNAPRIITVKIPVDKI 592

Qy        601 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 660
Db        593 GEVIGPKGKMINQIQEDTGAEITIEDDGTIYIGAQVGSQAEAARATINGIANPTMP---E 649

Qy        661 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQL 695
Db        650 VGERYLGTVVKTTTFGAFVSLLPGKDG-LLHISQI 683
```

Search completed: January 8, 2004, 10:42:19
Job time : 41 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 8, 2004, 10:33:56 ; Search time 27 Seconds
(without alignments)
2511.071 Million cell updates/sec

Title: US-09-907-907A-42
Perfect score: 3557
Sequence: 1 DGPFLLPRRDRALTQLQVRA.....TAVLLHNTQLDNERLNILLP 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1365.5	38.4	991	2	T48631	polynucleotide pho
2	1204.5	33.9	703	2	B97123	polyribonucleotide
3	1190.5	33.5	698	2	E89901	polyribonucleotide
4	1166.5	32.8	713	2	AE2586	polyribonucleotide
5	1166.5	32.8	713	2	F97368	polyribonucleotide
6	1161	32.6	711	2	AD0901	polynucleotide pho
7	1158.5	32.6	749	2	G97782	hypothetical prote
8	1158	32.6	734	2	E91134	polynucleotide pho
9	1158	32.6	734	2	H85979	polynucleotide pho
10	1156	32.5	810	2	G75320	polynucleotide pho
11	1153	32.4	718	2	S74509	polyribonucleotide
12	1152	32.4	734	2	H65106	polyribonucleotide
13	1150	32.3	714	2	AC3497	polyribonucleotide
14	1146	32.2	705	2	AC0424	polyribonucleotide
15	1145	32.2	704	2	G83950	polynucleotide pho
16	1140.5	32.1	745	2	B71654	polyribonucleotide
17	1140.5	32.1	775	2	E70320	polyribonucleotide
18	1138.5	32.0	718	2	AD2355	polyribonucleotide
19	1137	32.0	719	2	F82831	polynucleotide pho
20	1134	31.9	709	2	E64056	polyribonucleotide
21	1132	31.8	705	2	S70691	polyribonucleotide
22	1130	31.8	723	2	AG1603	polynucleotide pho
23	1127	31.7	723	2	AC1241	polynucleotide pho
24	1121	31.5	709	2	E82298	polyribonucleotide
25	1115	31.3	707	2	E84973	polyribonucleotide
26	1114.5	31.3	694	2	D86615	polyribonucleotide
27	1114.5	31.3	694	2	G72009	polyribonucleotide
28	1112	31.3	707	2	C81161	polyribonucleotide
29	1111	31.2	706	2	H81943	probable polyribon
30	1106	31.1	709	2	S38883	polyribonucleotide
31	1105	31.1	712	2	B87253	polyribonucleotide
32	1097	30.8	701	2	C83052	polyribonucleotide
33	1091	30.7	708	2	A72264	polynucleotide pho
34	1068.5	30.0	695	2	E71463	probable polyribon
35	1067.5	30.0	722	2	D70200	polyribonucleotide
36	1066	30.0	702	2	C71269	probable polyribon
37	1064	29.9	693	2	G81725	polyribonucleotide
38	1063.5	29.9	773	2	G86856	hypothetical prote
39	1054	29.6	737	2	D95068	polyribonucleotide
40	1053	29.6	775	2	D97936	polyribonucleotide
41	1029.5	28.9	719	2	G81332	polyribonucleotide
42	1016.5	28.6	897	2	T06540	polyribonucleotide
43	1005.5	28.3	739	2	T10932	polyribonucleotide
44	1000.5	28.1	777	2	T44900	probable polyribon
45	998.5	28.1	773	2	H87015	hypothetical prote

ALIGNMENTS

RESULT 1
T48631
polynucleotide phosphorylase - Arabidopsis thaliana
N;Alternate names: protein T15N1.70
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T48631
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24493
A;Accession: T48631
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-991 <BEV>
A;Cross-references: EMBL:AL163792
A;Experimental source: cultivar Columbia; BAC clone T15N1
C;Genetics:
A;Map position: 5
A;Introns: 107/3; 159/3; 216/3; 258/3; 312/3; 352/3; 398/3; 453/1; 495/1; 538/3; 612/1;
A;Note: T15N1.70

Query Match 38.4%; Score 1365.5; DB 2; Length 991;
Best Local Similarity 41.7%; Pred. No. 5.4e-78;
Matches 287; Conservative 138; Mismatches 231; Indels 33; Gaps 11;

```
Qy         24 SAGSRAV-----AVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQF 78
              |||:: :       ::|:| :   :||:||||:|| |:   :| |: |       ||  |
Db         44 SAGTKILESFKEEFEVGSRVVSFETGKIARFANGSVVLGMDETKVLSTVTCAKTDSPRDF 103

Qy         79 MPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCN 138
              :|| |||::|  | | ||  |:|||    ::|:|  |:||| |||||| |:::: |:: :
Db        104 LPLTVDYQEKQYAQGLIPNTYMRREGAPKERELLCGRLIDRPIRPLFPTGFYHEVQIMAS 163

Qy        139 LLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLN 198
              :|: ||  :||:|| | :| || |||:|| ||:| :||| | |::|||||  |:||| ||
Db        164 VLSSDGKQDPDILAANASSAALMLSDVPWGGPIGVIRIGRICGQFVVNPTMDELSSSDLN 223

Qy        199 LVVAGAPKSQIVMLEASAENILQQDFCHAIKV----GVKYTQQIIQGIQQLVKETGVTKR 254
              |: |   : : :|::  :  | ::|   |:::     |||    |    :| :: |  |:
Db        224 LIYA-CTRDKTMMIDVQSREISEKDLAAALRLAHPEAVKYLDPQI----RLAEKAGKQKK 278

Qy        255 TPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEA 314
                 ||   | : ::    ||  |: :|||  : |  | ||:: |  |  :  :|:  :
Db        279 -EYKLSMLSDKTLEKVADLAATRIESVFTDPSYGKFERGEALDNIGKDVRKVFEEEGDQE 337

Qy        315 DPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQT 374
                 : :: : | |:| || ::::  | ||| :  :| : ||      |||||| || |
Db        338 SLSILPKAVDTVRKKVVRSRMISDGFRVDGRHVDEVRPIYCESHYLPALHGSALFSRGDT 397

Qy        375 QVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHG 434
              |||||| :      | ::   |  | ||||| |||| |||:||  ||||||:|||
Db        398 QVLCTVTLGAPAEAQSLDSLV----GPPKKRFMLHYSFPPYCTNEVGKRGGLNRREVGHG 453
```

```
Qy        435 ALAEKALYPVIPRD--FPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGV 492
              ||||||  |:| :  ||:|||: |||: |:||:|||| ||||:||||:|:|: : ||||
Db        454 TLAEKALLAVLPPEEAFPYTIRINSEVMSSDGSTSMASVCGGSMALMDAGIPLRAHVAGV 513

Qy        493 AIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIK 552
              ::||:|  ||  |||:|||::|||||:||: |||||||||  |:||:| |||  |||:
Db        514 SVGLITDVDPSSGEIKDYRIVTDILGLEDHLGDMDFKIAGTRDGVTAIQLDIKPAGIPLD 573

Qy        553 IVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLK 612
              || |::: |  |: :||  | : |: ||       | : |::          :|| |   :
Db        574 IVCESLENAREARLQILDHMERNINSPRGQDGAYSPRLATLKYSNDSLRTLIGPMGVLKR 633

Qy        613 KLQAETGVTISQVDEETFSVFAPTPSVMHEAR---DFITEICKDDQEQQLEFGAVYTATI 669
              |:: |||  :| :|  | :: |    || :|:   |||          ::|  | ||  |:
Db        634 KIEVETGARLS-IDNGTLTIVAKNQDVMEKAQEQVDFII-------GRELVVGGVYKGTV 685

Qy        670 TEIRDTGVMVKLYPNMTAVLLHNTQLDNE 698
              : |:: |  |: :|     ||| ::| :|
Db        686 SSIKEYGAFVE-FPGGQQGLLHMSELSHE 713
```

RESULT 2
B97123
polyribonucleotide nucleotidyltransferase [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: B97123
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: B97123
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-703 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK79773.1; PID:g15024781; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1808
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

  Query Match 33.9%; Score 1204.5; DB 2; Length 703;
  Best Local Similarity 39.6%; Pred. No. 4.5e-68;
  Matches 266; Conservative 125; Mismatches 258; Indels 23; Gaps 10;

```
Qy         33 DLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPS-QFMPLVVDYRQKAAA 91
              |:  |||::  ||    :: :  :  ||| |||   :  ||     | || ::| ::  :
Db          8 DIAGRKLKVECGKTGMLSNCAMFISYGDTVVMVNVNASEKPREGIDFFPLSIEYEERQYS 67

Qy         92 AGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVL 151
              |:||  :::||   |:| || :| ||| :||||| ||  | ||:| :::|:  | |::|
Db         68 VGKIPGGFVKREGRPSEKSILHARAIDRPLRPLFPKGYRNDVQVVCTVMSVEQDNLPEIL 127

Qy        152 AINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVM 211
              |:||||:|| |||||:  ||  | :| |||::|:||| :|   |:|:| |  |   :::|
Db        128 AMNGASMALCLSDIPFTTPVATVSVGCIDGKFVLNPTLEEREKSSLDLTVC-ATNERVMM 186

Qy        212 LEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTH 271
              ||| |: | :     ||  |    | |:   :: :|| |  | ||: |: |  || |
Db        187 LEAGADEIPEDLMIAAIDFGFNACQDIVAFQEKAMKEFGKEKVTPE-LYHPKEEIEKDVT 245

Qy        272 KLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAKEVF 331
              : | :  :   |  |:  |:  : :|:     :  ||:|: |  :| |    : |:|
Db        246 EFAFESIKEIM--YITDRDERNLRLREIKEKISNEFAEKYPD-DGADIDEVVYTLQKKVV 302

Qy        332 RSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKS 391
              |:::| |::| |||    :| :||:||:    ||| || || |||:  |       |
Db        303 RNMLLKEHRRPDGRRFDEIRPISCDVDLLPRTHGSGLFTRGLTQVMTVTTL----GPIGD 358

Qy        392 DQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP--RDF 449
              |||  :   : | :| || ||||:| |:  : | ||||:|||||||||| |:||   :|
```

```
Db        359 AQVIDGLGVEESKRYMHHYNFPPYSTGEVKPLRGPNRREIGHGALAEKALVPLIPSEEEF 418

Qy        450 PFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEIED 509
              |:|||: |||| ||||:| || || :|||||:||||    ||:|:||:|  |  |  :
Db        419 PYTIRLVSEVLSSNGSTSQASVCGSTLALMDAGVPIKRPAAGIAMGLITSEDLSKEAV-- 476

Qy        510 YRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKEIL 569
                 :||| |:||: ||||||:||| |||||:| | |: |:    :  ||  |  |:  ||
Db        477 ---ITDIQGLEDFFGDMDFKVAGTEKGITAIQVDTKIHGLSKYCIKTAINDARKARLFIL 533

Qy        570 QIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDEET 629
              : |   |::||        |   |: :   |    :| ||  : |:  |||| |   :: |
Db        534 EKMVACINEPRKELSTYAPRAYTINIDTDKIRTLIGTGGKTINKIIEETGVKIDIREDGT 593

Qy        630 FSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMTAVL 689
              | :     : |   | :: ||     :: | ||   :|:| : |  |:: |      |
Db        594 VFVLSSDADSANRALKMIDDLTKD-----VKVGEVYLGKVTKITNFGAFVEVLPGKEG-L 647

Qy        690 LHNTQLDNERLN 701
              :| ::||  ::|
Db        648 VHISKLDINKVN 659
```

RESULT 3
E89901
polyribonucleotide nucleotidyltransferase [imported] - Staphylococcus aureus (strain N315
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: E89901
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguch
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; H
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: E89901
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-698 <KUR>
A;Cross-references: GB:BA000018; PID:g13701074; PIDN:BAB42369.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: pnpA
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

  Query Match 33.5%; Score 1190.5; DB 2; Length 698;
  Best Local Similarity 39.1%; Pred. No. 3.4e-67;
  Matches 273; Conservative 118; Mismatches 249; Indels 59; Gaps 12;

```
Qy         24 SAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVV 83
              |   :    :   | | | :|:||: |:|: :|: ||| |: || :  :|    | || |
Db          2 SQEKKVFKTEWAGRSLTIETGQLAKQANGAVLVRYGDTVVLSTATASKEPRDGDFFPLTV 61

Qy         84 DYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVD 143
              :| :|  |||:||  : :||    |   ||:|:||| |||||| || :| |::  :|: |
Db         62 NYEEKMYAAGKIPGGFKKREGRPGDDATLTARLIDRPIRPLFPKGYKHDVQIMNMVLSAD 121

Qy        144 GVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAG 203
              | : |: |:|:|||:|||| ||:  | :| |||:|::||| :|   | |:| |||
Db        122 PDCSPQMAAMIGSSMALSVSDIPFQGPIAGVNVGYIDGKYIINPTVEEKEVSRLDLEVAG 181

Qy        204 APKSQIVMLEASAENILQQDFCHAIKVG-------VKYTQQIIQGIQQLVKETGVTKRTP 256
              |  : |:|| |  | :|:   ||  |        | : |||:  || : :|
Db        182 -HKDAVNMVEAGASEITEQEMLEAIFFGHEEIQRLVDFQQQIVDHIQPVKQE-------- 232

Qy        257 QKLFTPS-------PEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKE 309
              | |:          :  | :  ::      |      ||  ||| :: ::   || : |
Db        233 ---FIPAERDEALVERVKSLTEEKGLKETVLTF-----DKQQRDENLDNLK---EEIVNE 281

Qy        310 KFPEADP------YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTL 363
              | ||       |:    | : ||  | :: :|  | |||    :| :  || :
Db        282 FIDEEDPENELLIKEVYAILNELVKEEVRRLIADEKIRPDGRKPDEIRPLDSEVGILPRT 341
```

```
Qy        364 HGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKV 423
Db        342 HGSGLFTRGQTQALSVLTLGAL----GDYQLIDGLGPEEEKRFMHHYNFPNFSVGETGPV 397

Qy        424 TGLNRRELGHGALAEKALYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDS 481
Db        398 RAPGRREIGHGALGERALKYIIPDTADFPYTIRIVSEVLESNGSSSQASICGSTLALMDA 457

Qy        482 GVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQ 541
Db        458 GVPIKAPVAGIAMGLVTRED-------SYTILTDIQGMEDALGDMDFKVAGTKEGITAIQ 510

Qy        542 ADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRA 601
Db        511 MDIKIDGLTREIIEEALEQARRGRLEIMNHMLQTIDQPRTELSAYAPKVVTMTIKPDKIR 570

Qy        602 KFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEF 661
Db        571 DVIGPGGKKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAREIIEEITR-----EAEV 625

Qy        662 GAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
Db        626 GQTYQATVKRIEKYGAFVGLFPGKDA-LLHISQISKNRI 663
```

RESULT 4
AE2586
polyribonucleotide nucleotidyltransferase pnpA [imported] - Agrobacterium tumefaciens (s
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AE2586
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AE2586
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-713 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL41107.1; PID:g17738399; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: pnpA
A;Map position: circular chromosome
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

  Query Match   32.8%; Score 1166.5; DB 2; Length 713;
  Best Local Similarity 39.1%; Pred. No. 1.2e-65;
  Matches 265; Conservative 133; Mismatches 251; Indels   29; Gaps   12;

```
Qy         29 AVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQ 87
Db          6 SVEIEWAGRPLKLETGKVARQADGAVIATYGETMVLATVVSAKSPKPGQDFFPLTVNYQE 65

Qy         88 KAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNE 147
Db         66 KTYAAGKIPGGYFKREGRPSEKETLVSRLIDRPIRPLFPEGYKNDTQVVVTVIQHDLEND 125

Qy        148 PDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKS 207
Db        126 PDVLSMVAASAALTLSGIPFMGPVGGARVGYINGEYVLNPHLDEMDESVLDLVVAGT-QD 184

Qy        208 QIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIV 267
Db        185 AVLMVESEAKELNEEIMLGAVMFGHKGFQPVIDAI---IKLAEVAAKEPRE-FEPEDFSA 240

Qy        268 KYTHKLAMERLYAVFTDYE-HDKVSRDEAVNKIRLDTEEQLKEKFPEA--DPYEIIESFN 324
Db        241 LENEMLGLAET-ELRTAYKITEKAARYAAVDAVKTKVKAHFLPEEGEAKYSPEEIGAVFK 299

Qy        325 VVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDS 384
Db        300 HLQAKIVRWNVLDTKSRIDGRDLSTVRPIVSEVGILPRTHGSALFTRGETQAIVVATLGT 359

Qy        385 LESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPV 444
Db        360 GE----DEQYVDSLTGMYKERFLLHYNFPPYSVGETGRMGSPGRREIGHGKLAWRAIRPM 415

Qy        445 IP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDP 502
Db        416 LPTAEQFPYTLRVVSEITESNGSSSMATVCGTSLALMDAGVPLAKPVAGIAMGLIL---- 471

Qy        503 EKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQAS 562
Db        472 ---EGERFAVLSDILGDEDHLGDMDFKVAGTADGITSLQMDIKIAGITEEIMKIALEQAQ 528

Qy        563 VAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTI 622
Db        529 GGRKHILGEMANAITESRGQLGEFAPRIEVMNIPVDKIREVIGSGGKVIREIVEKTGAKI 588

Qy        623 SQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLY 682
Db        589 NIEDDGTVKIASASGKEIEAARKWIHSIVAEP-----EVGQIYEGTVVKTADFGAFVNFF 643

Qy        683 PNMTAVLLHNTQLDNERL 700
Db        644 GARDG-LVHISQLASERV 660
```

RESULT 5
F97368
polyribonucleotide nucleotidyltransferase (PA4740) [imported] - Agrobacterium tumefacien
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: F97368
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: F97368
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-713 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK85903.1; PID:g15154950; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_124
A;Map position: circular chromosome
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

  Query Match   32.8%; Score 1166.5; DB 2; Length 713;
  Best Local Similarity 39.1%; Pred. No. 1.2e-65;
  Matches 265; Conservative 133; Mismatches 251; Indels   29; Gaps   12;

```
Qy         29 AVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQ 87
Db          6 SVEIEWAGRPLKLETGKVARQADGAVIATYGETMVLATVVSAKSPKPGQDFFPLTVNYQE 65

Qy         88 KAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNE 147
Db         66 KTYAAGKIPGGYFKREGRPSEKETLVSRLIDRPIRPLFPEGYKNDTQVVVTVIQHDLEND 125

Qy        148 PDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKS 207
Db        126 PDVLSMVAASAALTLSGIPFMGPVGGARVGYINGEYVLNPHLDEMDESVLDLVVAGT-QD 184

Qy        208 QIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIV 267
Db        185 AVLMVESEAKELNEEIMLGAVMFGHKGFQPVIDAI---IKLAEVAAKEPRE-FEPEDFSA 240

Qy        268 KYTHKLAMERLYAVFTDYE-HDKVSRDEAVNKIRLDTEEQLKEKFPEA--DPYEIIESFN 324
```

```
Db        241 LENEMLGLAET-ELRTAYKITEKAARYAAVDAVKTKVKAHFLPEEGEAKYSPEEIGAVFK 299

Qy        325 VVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDS 384

Db        300 HLQAKIVRWNVLDTKSRIDGRDLSTVRPIVSEVGILPRTHGSALFTRGETQAIVVATLGT 359

Qy        385 LESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPV 444

Db        360 GE----DEQYVDSLTGMYKERFLLHYNFPPYSVGETGRMGSPGRREIGHGKLAWRAIRPM 415

Qy        445 IP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDP 502

Db        416 LPTAEQFPYTLRVVSEITESNGSSSMATVCGTSLALMDAGVPLAKPVAGIAMGLIL---- 471

Qy        503 EKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQAS 562

Db        472 ---EGERFAVLSDILGDEDHLGDMDFKVAGTADGITSLQMDIKIAGITEEIMKIALEQAQ 528

Qy        563 VAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTI 622

Db        529 GGRKHILGEMANAITESRGQLGEFAPRIEVMNIPVDKIREVIGSGGKVIREIVEKTGAKI 588

Qy        623 SQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLY 682

Db        589 NIEDDGTVKIASASGKEIEAARKWIHSIVAEP-----EVGQIYEGTVVKTADFGAFVNFF 643

Qy        683 PNMTAVLLHNTQLDNERL 700

Db        644 GARDG-LVHISQLASERV 660


RESULT 6
AD0901
polynucleotide phosphorylase [imported] - Salmonella enterica subsp. enterica serovar Ty
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AD0901
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AD0901
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-711 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD07802.1; PID:g16504350; GSPDB:GN00176
C;Genetics:
A;Gene: STY3463
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

  Query Match             32.6%;  Score 1161;  DB 2;  Length 711;
  Best Local Similarity   39.9%;  Pred. No. 2.6e-65;
  Matches  270;  Conservative  118;  Mismatches  249;  Indels   40;  Gaps   11;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93

Db         12 GQHTVTLETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQDFFPLTVNYQERTYAAG 71

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153

Db         72 RIPGSFFRREGRPSEGETLIARLIDRPVRPLFPEGFVNEVQVIATVVSVNPQVNPDIVAM 131

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213

Db        132 IGASAALSLSGIPFNGPIGAARVGYINDQYVLNPTQDELKESKLDLVVAGT-EAAVLMVE 190

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVK----- 268

Db        191 SEAELLSEDTMLGAVVFGHEQQQVVIQAINDLVKEAGKPRWDWQ------PEAVNDALNA 244
```

```
Qy        269 YTHKLAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNV 325

Db        245 RVAALAESRLSDAYRI-----TDKQERYAQVDVIKSETIEQLIAEDETLDANELGEILHA 299

Qy        326 VAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSL 385

Db        300 IEKNVVRSRVLAGEPRIDGREKDMIRGLDVRTGVLPRTHGSALFTRGETQALVTATLGT- 358

Qy        386 ESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVI 445

Db        359 ---ARDAQVLDELMGERTDSFLFHYNFPPYSVGETGMVGSPKRREIGHGRLAKRGVLAVM 415

Qy        446 P--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPE 503

Db        416 PDMDKFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVKEGD-- 473

Qy        504 KGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASV 563

Db        474 -----NYVVLSDILGDEDHLGDMDFKVAGSRDGISALQMDIKIEGITKEIMQVALNQAKG 528

Qy        564 AKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTIS 623

Db        529 ARLHILGVMEQAINAPRGDISEFAPRIHTIKISTDKIKDVIGKGGSVIRALTEETGTTIE 588

Qy        624 QVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYP 683

Db        589 IEDDGTVKIAATDGEKAKYAIRRIEEI-----TAEIEVGRIYNSKVTRIVDFGAFVAIGG 643

Qy        684 NMTAVLLHNTQLDNERL 700

Db        644 GKEG-LVHISQIADKRV 659


RESULT 7
G97782
hypothetical protein pnp [imported] - Rickettsia conorii (strain Malish 7)
C;Species: Rickettsia conorii
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C;Accession: G97782
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Rou
Science 293, 2093-2098, 2001
A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A;Reference number: A97700; MUID:21442074; PMID:11557893
A;Accession: G97782
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-749 <KUR>
A;Cross-references: GB:AE006914; PIDN:AAL03201.1; PID:g15619751; GSPDB:GN00173
C;Genetics:
A;Gene: pnp
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

  Query Match             32.6%;  Score 1158.5;  DB 2;  Length 749;
  Best Local Similarity   38.0%;  Pred. No. 4e-65;
  Matches  257;  Conservative  145;  Mismatches  244;  Indels   31;  Gaps   12;

Qy         31 AVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPS-QFMPLVVDYRQKA 89

Db          8 SVTWNGQVLELSTGKIARQADGAVTVKMGNSVLLCTAVVANKAKEGIGFLPLTINYREMA 67

Qy         90 AAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPD 149

Db         68 YAAGKIPGGFFKHEGKASDREVLVSRLIDRPIRPLFHPAFVNETHVTCSVLSYDPETPVD 127

Qy        150 VLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQI 209

Db        128 ILAIIGASAALSLSPAPYLEIVAASKVGLINGEFVLNPTLALLKTSQLDLVVAGTSDS-V 186

Qy        210 VMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKY 269

Db        187 MMVESEAHLLSEEQMLEAVKFGFESFQPVIKIIKELAEEAKKPKLEMQALYPAS--LKKE 244
```

```
Qy        270 THKL---AMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVV 326
Db        245 IEKLFVKEIEQAFAIKSKQER-STNLDLIPEKVLTHFVSDIENK--KYSNYQIESALKAI 301

Qy        327 AKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTF-DSL 385
Db        302 ESDILRNEILEKNRRIDGRSTTDIRQIACEIGLLPSAHGSALFTRGETQSLVSTTFGTSL 361

Qy        386 ESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVI 445
Db        362 D-----EQIVDSLEGEYKERFMLNYIFPPYSVNEAMPMKAPSRREVGHGKLAWRAINPIL 416

Qy        446 PR--DFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPE 503
Db        417 PNKVQFPYSIRVVAETTESNGSSSMATVCGSSLALMYAGVPIKAPVAGIAMGLVK----- 471

Qy        504 KGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASV 563
Db        472 --EGKNFAVLSDILGDEDYFGDMDFKVAGTSEGITALQMDIKISGVDFKIMKVALEQARL 529

Qy        564 AKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTIS 623
Db        530 GRLHILEQMNKVISKPNNELSKNAPSTTTIKIDKDKIRDIIGPGGKIIKEICETSGAKID 589

Qy        624 QVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYP 683
Db        590 ISDDGTVSVYASDRDKLKVALDKIKAIVVEP-----EIGEIFNGTVVKVLDSGAFINYVG 644

Qy        684 NMTAVLLHNTQLDNERL 700
Db        645 NKDG-FVHISEVSGERI 660
```

RESULT 8
E91134
polynucleotide phosphorylase [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: E91134
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E91134
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-734 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB37468.1; PID:g13363518; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4045
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

Query Match 32.6%; Score 1158; DB 2; Length 734;
Best Local Similarity 40.0%; Pred. No. 4.2e-65;
Matches 270; Conservative 117; Mismatches 252; Indels 36; Gaps 11;

```
Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
Db         35 GQHTVTLETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQDFFPLTVNYQERTYAAG 94

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
Db         95 RIPGSFFRREGRPSEGETLIARLIDRPIRPLFPEGFVNEVQVIATVVSVNPQVNPDIVAM 154

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
Db        155 IGASAALSLSGIPFNGPIGAARVGYINDQYVLNPTQDELKESKLDLVVAGT-EAAVLMVE 213

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSP---EIVKYT 270
Db        214 SEAELLSEDQMLGAVVFGHEQQQVVIQNINELVKEAG----KPRWDWQPEPVNEALNARV 269
```

```
Qy        271 HKLAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVA 327
Db        270 AALAEARLSDAYRI-----TDKQERYAQVDVIKSETIATLLAEDETLDENELGEILHAIE 324

Qy        328 KEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLES 387
Db        325 KNVVRSRVLAGEPRIDGREKDMIRGLDVRTGVLPRTHGSALFTRGETQALVTATLGT--- 381

Qy        388 GIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP- 446
Db        382 -ARDAQVLDELMGERTDTPLFHYNFPPYSVGETGMVGSPKRREIGHGRLAKRGVLAVMPD 440

Qy        447 -RDFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKG 505
Db        441 MDKFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVKEGD---- 496

Qy        506 EIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAK 565
Db        497 ---NYVVLSDILGDEDHLGDMDFKVAGSRDGISALQMDIKIEGITKEIMQVALNQAKGAR 553

Qy        566 KEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQV 625
Db        554 LHILGVMEQAINAPRGDISEFAPRIHTIKINPDKIKDVIGKGGSVIRALTEETGTTIEIE 613

Qy        626 DEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNM 685
Db        614 DDGTVKIAATDGEKAKHAIRRIEEI-----TAEIEVGRVYTGKVTRIVDFGAFVAIGGGK 668

Qy        686 TAVLLHNTQLDNERL 700
Db        669 EG-LVHISQIADKRV 682
```

RESULT 9
H85979
polynucleotide phosphorylase [imported] - Escherichia coli (strain O157:H7, substrain ED
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001
C;Accession: H85979
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: H85979
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-734 <STO>
A;Cross-references: GB:AE005174; NID:g12517775; PIDN:AAG58300.1; GSPDB:GN00145; UWGP:Z45
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: pnp
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

Query Match 32.6%; Score 1158; DB 2; Length 734;
Best Local Similarity 40.0%; Pred. No. 4.2e-65;
Matches 270; Conservative 117; Mismatches 252; Indels 36; Gaps 11;

```
Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
Db         35 GQHTVTLETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQDFFPLTVNYQERTYAAG 94

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
Db         95 RIPGSFFRREGRPSEGETLIARLIDRPIRPLFPEGFVNEVQVIATVVSVNPQVNPDIVAM 154

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
Db        155 IGASAALSLSGIPFNGPIGAARVGYINDQYVLNPTQDELKESKLDLVVAGT-EAAVLMVE 213

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSP---EIVKYT 270
Db        214 SEAELLSEDQMLGAVVFGHEQQQVVIQNINELVKEAG----KPRWDWQPEPVNEALNARV 269
```

```
Db        214 SEAELLSEDQMLGAVVFGHEQQQVVIQNINELVKEAG----KPRWDWQPEPVNEALNARV 269

Qy        271 HKLAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVA 327

Db        270 AALAEARLSDAYRI-----TDKQERYAQVDVIKSETIATLLAEDETLDENELGEILHAIE 324

Qy        328 KEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLES 387

Db        325 KNVVRSRVLAGEPRIDGREKDMIRGLDVRTGVLPRTHGSALFTRGETQALVTATLGT--- 381

Qy        388 GIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP- 446

Db        382 -ARDAQVLDELMGERTDTFLFHYNFPPYSVGETGMVGSPKRREIGHGRLAKRGVLAVMPD 440

Qy        447 -RDFPPTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKG 505

Db        441 MDKFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVKEGD---- 496

Qy        506 EIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAK 565

Db        497 ---NYVVLSDILGDEDHLGDMDFKVAGSRDGISALQMDIKIEGITKEIMQVALNQAKGAR 553

Qy        566 KEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQV 625

Db        554 LHILGVMEQAINAPRGDISEFAPRIHTIKINPDKIKDVIGKGGSVIRALTEETGTTIEIE 613

Qy        626 DEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNM 685

Db        614 DDGTVKIAATDGEKAKHAIRRIEEI-----TAEIEVGRVYTGKVTRIVDFGAFVAIGGGK 668

Qy        686 TAVLLHNTQLDNERL 700

Db        669 EG-LVHISQIADKRV 682


RESULT 10
G75320
polynucleotide phosphorylase - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: G75320
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: G75320
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-810 <WHI>
A;Cross-references: GB:AE002042; GB:AE000513; NID:g6459848; PIDN:AAF11608.1; PID:g645985
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2063
A;Map position: 1
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

  Query Match             32.5%;  Score 1156;  DB 2;  Length 810;
  Best Local Similarity   38.1%;  Pred. No. 6.5e-65;
  Matches  261;  Conservative  138;  Mismatches  236;  Indels   50;  Gaps   14;

Qy         34 LGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAG 93

Db         41 LGGRELSIETGKLAKLVSGSVTVRYGDTLLLVTAQASDTQSKLDFLPLTVEPEERHYAVG 100

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153

Db        101 KIPGSFQRREGRPGEKAILSARITDRQIRPLFPKGYRHETQVIITVLSADGQNAPDVLGP 160

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213

Db        161 IGAAAALSISDIPWAGPTACVRVGQIDGQYVVNPTTEQLTRSRMDLVVAGT-REAVMMVE 219

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETG------VTKRTPQKLFTPSPEIV 267

Db        220 CGAQTVSEDDLVGAIEFAHAEMQGVIALIEQMRAEVGHEKFNFLAEEGPANDYVP----- 274

Qy        268 KYTHKLAMERLY-AVFTDYEHDKVSRDEAVNKIRL-----DTEEQLKEKFPEADPYEIIE 321

Db        275 ELTEKAKAAGLRDALLTHGKKDRSARLKALRNGLIEGYVPDPTAEGSAELTQA----LKD 330

Qy        322 SFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVT 381

Db        331 AFGKVEKRELRRLILEENLRADGRDSKTVRPIWIEARPLPTAHGSAVFTRGETQVLGVTT 390

Qy        382 FDSLESGIKSDQV----ITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALA 437

Db        391 L-----GTERDEILIDDLTAESGDK---FLLHYNFPPYSTGEVKRMGGQSRREIGHGNLA 442

Qy        438 EKALYPVIP--RDFPPTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIG 495

Db        443 KRAIRAVLPSFEEFPYVIRVVGDVLESNGSSSMGTVCAGTLSLMDAGVPLKAPVAGVAMG 502

Qy        496 LVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVM 555

Db        503 LVMEGD-------NYRVLTDILGLEDALGDMDFKVCGTAEGVTALQMDIKVGGITPQIMR 555

Qy        556 EAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQ 615

Db        556 EALAQAKEGRLHILGKMAEVLAAPRAELSPTAPHILSLKINPELIGKVIGPGGKQVRELE 615

Qy        616 AETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDT 675

Db        616 A-MGAQVTIEEDGTVRIFSASGESAEAVKARIEAVTK-----EAKVGEEFEGTVVKIAPF 669

Qy        676 GVMVKLYPNMTAVLLHNTQLDNERL 700

Db        670 GAFVNLFPGQDG-MLHISQLSEQRV 693


RESULT 11
S74509
polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) alpha chain - Synechocystis sp. (s
N;Alternate names: protein sll1043
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S74509
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S74509
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-718 <KAN>
A;Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PIDN:BAA16661.1; PID:g165173
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Gene: pnp
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain
C;Keywords: nucleotidyltransferase

  Query Match             32.4%;  Score 1153;  DB 2;  Length 718;
  Best Local Similarity   37.5%;  Pred. No. 8.3e-65;
  Matches  257;  Conservative  131;  Mismatches  261;  Indels   36;  Gaps   10;

Qy         31 AVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAA 90

Db          7 SISFDGRDIRLKMGTLAPQAGGSVLIQSGDTAVLVTATRAKGRDGIDFLPLTVDYEGRLY 66

Qy         91 AAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDV 150
```

```
Db         67 AAGRIPGGFLRREGRPPEKATLISRLIDRPLRPLFPHWLRDELQIVATTLSMDEEVPPDV 126

Qy        151 LAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIV 210

Db        127 LAVTGASVAVILAQIPFKGPMAAAVRVGLVGDDFIINPTYREVHNGDLDLVVAGTP-AGIV 185

Qy        211 MLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSP---EIV 267

Db        186 MVEAGANQLPEQDIIEAIDFGYEAVQDLINAQRELMTDLGITLATSE----PPPVNTAVE 241

Qy        268 KYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYE--------- 318

Db        242 EFIANRASKKIITVLGQFDLGKDGRDAALDEIKATEVETAIAELPETDPVKQSVEEDPKL 301

Qy        319 IIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF-KTLHGSALFQRGQTQVL 377

Db        302 VGNLYKALTKKLMRKQIVDDGVRVDGRKLEQVRPISCEVGFLPRRVHGSGLFNRGLTQVL 361

Qy        378 CTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALA 437

Db        362 SLAT---LGSPGDAQDLADDLHPEDEKRYLHHYNFPPYSVGEARPMRSPGRREIGHGALA 418

Qy        438 EKALYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIG 495

Db        419 ERAIIPVLPPQEDFPYVVRVVSEVLSSNGSTSMGSVCGSTLALMDAGVPIKKPVSGAAMG 478

Qy        496 LVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVM 555

Db        479 LIKEGD-------EIRILTDIQGIEDFLGDMDFKVAGTDSGITALQMDMKIDGLSMEVVS 531

Qy        556 EAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQ 615

Db        532 KAIMQALPARLHILDKMLATIREPRPELSPFAPRLLTLKIEPEHIGMVIGPGGKTIKGIT 591

Qy        616 AETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDT 675

Db        592 EQTSCKIDIADDGTVTIASSEGERAERARQMIYNMTR-----KLNEGEVYLGRVTRIIPI 646

Qy        676 GVMVKLYPNMTAVLLHNTQLDNERL 700

Db        647 GAFVEVLPGKEG-MIHISQLTEGRV 670


RESULT 12
H65106
polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) alpha chain - Escherichia coli (s
N;Alternate names: polynucleotide phosphorylase
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 23-Jan-1998 #text_change 01-Mar-2002
C;Accession: H65106; A26118; B23984
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H65106
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-734 <BLAT>
A;Cross-references: GB:AE000397; GB:U00096; NID:g2367199; PIDN:AAC76198.1; PID:g1789555;
A;Experimental source: strain K-12, substrain MG1655
R;Regnier, P.; Grunberg-Manago, M.; Portier, C.
J. Biol. Chem. 262, 63-68, 1987
A;Title: Nucleotide sequence of the pnp gene of Escherichia coli encoding polynucleotide
in S1.
A;Reference number: A26118; MUID:87083499; PMID:2432069
A;Accession: A26118
A;Molecule type: DNA
A;Residues: 'M',25-379,'R',381-472,'S',474-734 <REG>
A;Cross-references: EMBL:J02638; NID:g147744; PIDN:AAA83905.1; PID:g551833
R;Evans, S.; Dennis, P.P.
Gene 40, 15-22, 1985
A;Title: Promoter activity and transcript mapping in the regulatory region for genes end
A;Reference number: A23984; MUID:86137413; PMID:3005122
A;Accession: B23984
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 'M',25-85 <EVA>
A;Cross-references: GB:M14425; NID:g147747; PIDN:AAA24596.1; PID:g147749
A;Note: the authors translated the codon GCC for residue 74 as Arg
C;Genetics:
A;Gene: pnp
A;Map position: 69 min
A;Start codon: TTG
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain
C;Keywords: nucleotidyltransferase
F;25-734/Product: polyribonucleotide nucleotidyltransferase alpha chain #status predicte

  Query Match            32.4%;  Score 1152;  DB 2;  Length 734;
  Best Local Similarity  39.9%;  Pred. No. 1e-64;
  Matches  269;  Conservative  118;  Mismatches  252;  Indels   36;  Gaps   11;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93

Db         35 GQHTVTLETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQDFFPLTVNYQERTYAAG 94

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153

Db         95 RIPGSFFRREGRPSEGETLIARLIDRPIRPLFPEGFVNEVQVIATVVSVNPQVNPDIVAM 154

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213

Db        155 IGASAALSLSGIPFNGPIGAARVGYINDQYVLNPTQDELKESKLDLVVAGT-EAAVLMVE 213

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSP---EIVKYT 270

Db        214 SEAQLLSEDQMLGAVVFGHEQQQVVIQNINELVKEAG----KPRWDWQPEPVNEALNARV 269

Qy        271 HKLAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVA 327

Db        270 AALAEARLSDAYRI-----TDKQERYAQVDVIKSETIATLLAEDETLDENELGEILHAIE 324

Qy        328 KEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLES 387

Db        325 KNVVRSRVLAGEPRIDGREKDMIRGLDVRTGVLPRTHGSALFTRGETQALVTATLGT--- 381

Qy        388 GIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP- 446

Db        382 -ARDAQVLDELMGERTDTFLFHYNFPPYSVGETGMVGSPKRREIGHGRLAKRGVLAVMPD 440

Qy        447 -RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKG 505

Db        441 MDKFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVKEGD---- 496

Qy        506 EIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAK 565

Db        497 ---NYVVLSDILGDEDHLGDMDFKVAGSRDGISALQMDIKIEGITKEIMQVALNQAKGAR 553

Qy        566 KEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQV 625

Db        554 LIILGVMEQAINAPRGDISEFAPRIHTIKINPDKIKDVIGKGGSVIRALTEETGTTIEIE 613

Qy        626 DEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNM 685

Db        614 DDGTVKIAATDGEKAKHAIRRIEEI-----TAEIEVGRVYTGKVTRIVDFGAFVAIGGGK 668

Qy        686 TAVLLHNTQLDNERL 700

Db        669 EG-LVHISQIADKRV 682


RESULT 13
AC3497
polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) [imported] - Brucella melitensis
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
```

```
C;Accession: AC3497
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letese
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AC3497
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-714 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL53142.1; PID:g17984011; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI1961
A;Map position: I
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain
C;Keywords: nucleotidyltransferase

  Query Match             32.3%;  Score 1150;  DB 2;  Length 714;
  Best Local Similarity   39.0%;  Pred. No. 1.3e-64;
  Matches  268;  Conservative  124;  Mismatches  247;  Indels   48;  Gaps   14;

Qy         30 VAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQK 88
              | :: | | | : :||:|| |||: :   |:|||: | ||  :| | | | || |:|::|
Db          7 VEIEWGGRPLTLETGKIARQADGAVLATYGETAVLATVVSAKEPKPGQDFFPLTVNYQEK 66

Qy         89 AAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEP 148
                 |||:|| | :||   |: | | ||:||| |||||  ||  ||||:  :|  |  | |
Db         67 TYAAGKIPGGYFKREGRPSENETLVSRLIDRPIRPLFVDGYKNDTQVVITVLQHDLENNP 126

Qy        149 DVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQ 208
              |:|::  || ||::| :|: ||:   |:| ||||||:||   ||  | |:|||||  ::
Db        127 DILSMVAASAALTISGVPFMGPISGARVGYIDGEYVLNPNIDEMPESKLDLVVAGTSEA- 185

Qy        209 IVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVK 268
              ::|:|: |: : :     |:  | |  | :|  |   :|   |  : |:      || :
Db        186 VLMVESEAQELPEDVMLGAVMFGHKSFQPVIDAI---IKLAEVAAKEPRDF---QPEDLS 239

Qy        269 YTHKLAMERLYAVFTDYEH-----DKVSRDEAVNKIRLDTEEQLKEKFPE--------AD 315
                :|  : |  |  |          :| :|  ||:  :   ||    |||          |:
Db        240 ---ELEAKVLAVVENDLREAYKITEKQARYAAVDAAKAKAEEHF---FPEGVEETEMSAE 293

Qy        316 PYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQ 375
              :  |  |  :  :: |  :|:    | |||||:::| :  || :    ||||| ||:||
Db        294 QFATI--FKHLQAKIVRWNILDTGNRIDGRDLSTVRPIVSEVGILPRTHGSALFTRGETQ 351

Qy        376 VLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGA 435
              :   |  : |      :|:|  |: |   ::|||| ||||:  | |::    |||:|||
Db        352 AIVVATLGTGE----DEQMIDALTGTYKESFMLHYNFPPYSVGETGRMGSPGRREIGHGK 407

Qy        436 LAEKALYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVA 493
              || :|::|::|    ||:||| ||: |||||||||: || ||||||:||||   |||:|
Db        408 LAWRAIHPMLPAAEQFPYTIRAVSEITESNGSSSMATVCGTSLALMDAGVPIVRPVAGIA 467

Qy        494 IGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKI 553
              :||:        | | : :|:|||| ||: |||||:|||  |||:|| |||: ||  :|
Db        468 MGLIK-------EGERFAVLSDILGDEDHLGDMDFKVAGTEFGITSLQMDIKIDGITEEI 520

Qy        554 VMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKK 613
              :  |::||    :  ||  | | ||  ||    |  | :| : :|  |     :| ||  :::
Db        521 MKVALEQAKGGRVHILGEMAKAISSSRAELGEFAPRIEVMNIPTDKIRDVIGSGGKVIRE 580

Qy        614 LQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIR 673
              :  :||  |:  |: |  : :     :  |: :|  |  :      | | :|  |: :
Db        581 IVEKTGAKINIEDDGTVKIASSNGKEIEAAKKWIHSIVAEP-----EVGEIYEGTVVKTA 635

Qy        674 DTGVMVKLYPNMTAVLLHNTQLDNERL 700
              | |  |  :      |:| :||  :|:
Db        636 DFGAFVNFFGPRDG-LVHISQLAADRV 661


RESULT 14
AC0424
```

```
polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) [imported] - Yersinia pestis (stra
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C;Accession: AC0424
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.,
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; E
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, E
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0424
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-705 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92719.1; PID:g15981414; GSPDB:GN00175
C;Genetics:
A;Gene: pnp
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain
C;Keywords: nucleotidyltransferase

  Query Match             32.2%;  Score 1146;  DB 2;  Length 705;
  Best Local Similarity   40.0%;  Pred. No. 2.2e-64;
  Matches  269;  Conservative  113;  Mismatches  259;  Indels   32;  Gaps   10;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
              |   : | :| :|| |  : :|   |||| || | : |  | | | || |:|::  |||
Db         12 GQHTVTIETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQSFFPLTVNYQERTYAAG 71

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
              ||| :: |||    |: | ||||:||| |||||  :  : ||:  :::|:    ||::|:
Db         72 RIPGSFFRREGRPSEGETLTSRLIDRPIRPLFPDSFLNEVQVIATVVSVNPQINPDIVAL 131

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
               ||| ||||| ||:|||:|| |:| |: :||:|||  |:  | |:|||||      ::|:|
Db        132 IGASAALSLSGIPFNGPIGAARVGFINDQYVLNPTTDELKESRLDLVVAGT-AGAVLMVE 190

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTH-- 271
              : |: : ::    |:  | :  | :|: |  || | |  |    |      :  : :  |
Db        191 SEADILSEEQMLGAVVFGHEQQQVVIENINALVAEAGKPKWDWQ-----AEPVNEALHAR 245

Qy        272 --KLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAKE 329
                :||  ||   :   |  |  |   |: |: |  | |  :    |  || :     | |
Db        246 VAELAEARLGDAYRITE--KQERYTQVDAIKADVTEALLAQDDTLDAAEIQDILASVEKN 303

Qy        330 VFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGI 389
              | || ||    | |||:   :| :    :    |||||| ||:|| | | |  :
Db        304 VVRSRVLRGEPRIDGREKDMIRGLDVRTGILPRTHGSALFTRGETQALVTATLGTARDAQ 363

Qy        390 KSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP--R 447
              |:::    | :  :|:||| ||||   | | |    |||:||| ||:: :  |:|
Db        364 NIDELM----GERTDSFLLHYNFPPYCVGETGMVGSPKRREIGHGRLAKRGVLAVMPSAS 419

Qy        448 DFPPTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEI 507
              :||:|||| ||: |||||||||| || ||||||:|||| :||||:|:||| : |
Db        420 EFPPYTIRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVKEGD------ 473

Qy        508 EDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKE 567
              :: :|:|||| ||: ||||||:||:  |:|||| |||: ||  :|:  |: ||  |:
Db        474 -NFVVLSDILGDEDHLGDMDFKVAGSRDGVTALQMDIKIEGITREIMQVALNQAKGARLH 532

Qy        568 ILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDE 627
              || :| : || ||    |  | : |:::    |     :| ||  :: |  ||| ||   |:
Db        533 ILGVMEQAISTPRGDISEFAPRIYTMKINPEKIKDVIGKGGSVIRALTDETGTTIEIEDD 592

Qy        628 ETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMTA 687
              |  : |         |    | ||         ::| | :|   :| | | | | :
Db        593 GTIKIAATDGDKAKHAIRRIEEI-----TAEIEVGRIYAGKVTRIVDFGAFVAIGGGKEG 647

Qy        688 VLLHNTQLDNERL 700
              |:| :|: ::|:
Db        648 -LVHISQIADKRV 659
```

```
RESULT 15
G83950
polynucleotide phosphorylase pnpA [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: G83950
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: G83950
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-704 <STO>
A;Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BAB06126.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: pnpA
C;Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

  Query Match             32.2%;  Score 1145;  DB 2;  Length 704;
  Best Local Similarity   39.3%;  Pred. No. 2.6e-64;
  Matches  265;  Conservative  132;  Mismatches  251;  Indels   26;  Gaps   11;

Qy         31 AVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAA 90
              ::|   ||| : :|:||: |:|: :|: |||||: || :  :|    | || |:| ::
Db          9 SIDWAGRKLTVETGQLAKQANGAVLVRYGDTAVLSTATASKEPKDLPFFPLTVNYEERLY 68

Qy         91 AAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDV 150
              |||:||  :::||  |:| || ||:||| |||||| |:  : ||:  :::||     ::
Db         69 AAGKIPGGFIKREGRPSEKAILASRLIDRPIRPLFPEGFRNEVQVISIVMSVDQDCSSEM 128

Qy        151 LAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIV 210
              |: |:|:|||:|||: ||:  | :| |||::|:|||: ::  | ::|||||  |  |
Db        129 AAMVGSSLALSISDIPFEGPIAGVTVGRIDGQFVINPTQDQLEKSDIHLVVAGT-KDAIN 187

Qy        211 MLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYT 270
              |:|| || : :     ||  |    :::|:  :::  | | :| |   |    | : :
Db        188 MVEAGAEEVPEDVMLEAIMFGHNEIKRLIEFQEKIAAEVGKSK-TDVVLKQVDPMLEQEV 246

Qy        271 HKLAMERL-YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAKE 329
                | | |  ||    :| :    |||    |:| |   |  | ::  ||:    : ||
Db        247 RVKAEEDLKQAVQVPEKHARQDAIEAVMDKVLETYED-NEDVPLSEVNEILHK---IVKE 302

Qy        330 VFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGI 389
              | ::  |  | |||::  :| :| :| :    ||| || ||||| |   |  :|
Db        303 EVRRLITVEKIRPDGREIDEIRPLSSQVGILPRTHGSGLFTRGQTQALSICTLGAL---- 358

Qy        390 KSDQVITAINGIKDKNFMLHYEPPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP--R 447
              |::  :   : | || || || ::  | | : |  |||:|||| |:|| ||||  :
Db        359 GDVQILDGLGIEESKRFMHHYNFPQFSVGETGPIRGPGRREIGHGALGERALEPVIPSEQ 418

Qy        448 DFPPTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEI 507
              |||:|||: ||||||||||:| || |  :||:||:|||| : |||:|:||| | |
Db        419 DFPYTIRLVSEVLESNGSTSQASICASTLAMMDAGVPIKAPVAGIAMGLV-KQD------ 471

Qy        508 EDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKE 567
              |   :|||| |:||  ||||||:||| ||:|||| |||: ||   |: :|::||   :
Db        472 EHVSVLTDIQGMEDALGDMDFKVAGTRKGVTALQMDIKISGIDRAILEQALEQARKGRMI 531

Qy        568 ILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDE 627
              ||  | : ||: |:       | : |: :   |    :|| |  : |:  :||| |    :
Db        532 ILDNMLEAISESRSELSPYAPKILTMTINPDKIRDVIGPSGKMINKIIEDTGVKIDIEQD 591

Qy        628 ETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMTA 687
              |  : :   :: ::||: | :| :     ::| | :|  |:  |   |  |:|:
Db        592 GTIYISSADTNMNNKAREIIEDIVR-----EVEVGQMYLGTVKRIEKFGAFVELFKGKDG 646

Qy        688 VLLHNTQLDNERLN 701
              |:| :||  ||:|
Db        647 -LVHISQLAEERVN 659
```

```
Search completed: January  8, 2004, 10:40:52
Job time : 29 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 8, 2004, 10:30:06 ; Search time 17 Seconds
(without alignments)
1950.225 Million cell updates/sec

Title: US-09-907-907A-42
Perfect score: 3557
Sequence: 1 DGPFLLPRRDRALTQLQVRA.....TAVLLHNTQLDNERLNILLP 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1155	32.5	711	1	PNP_ECOLI	P05055 escherichia
2	1152	32.4	707	1	PNP_BUCAP	Q8k9h5 buchnera ap
3	1140.5	32.1	745	1	PNP_RICPR	Q9zd43 rickettsia
4	1136	31.9	706	1	PNP_YEREN	O34275 yersinia en
5	1134	31.9	709	1	PNP_HAEIN	P44584 haemophilus
6	1132	31.8	704	1	PNP_BACSU	P50849 bacillus su
7	1115	31.3	707	1	PNP_BUCAI	P57454 buchnera ap
8	1106	31.1	709	1	PNP_PHOLU	P41121 photorhabdu
9	1068	30.0	701	1	PNP_PSEPU	O87792 pseudomonas
10	209	5.9	248	1	ECX1_SULSO	Q9uxc2 sulfolobus
11	208.5	5.9	246	1	ECX1_AERPE	Q9yc03 aeropyrum p
12	193.5	5.4	493	1	ECX1_METMA	Q8ptt8 methanosarc
13	192	5.4	247	1	ECX1_SULTO	Q975g8 sulfolobus
14	186	5.2	246	1	ECX1_PYRAE	Q8zvm9 pyrobaculum
15	175.5	4.9	249	1	ECX1_PYRAB	Q9v119 pyrococcus
16	175	4.9	258	1	ECX1_ARCFU	O29757 archaeoglob
17	173.5	4.9	249	1	ECX1_PYRHO	O59223 pyrococcus
18	168.5	4.7	250	1	ECX1_PYRFU	Q8u019 pyrococcus
19	163	4.6	261	1	RNPH_THETN	Q8rc30 thermoanaer
20	157.5	4.4	255	1	RNPH_AQUAE	O67069 aquifex aeo
21	156.5	4.4	245	1	RNPH_BACSU	P28619 bacillus su
22	155	4.4	248	1	RNPH_LISMO	Q8y7n6 listeria mo
23	151	4.2	257	1	RNPH_BACHD	Q9k8d8 bacillus ha
24	150.5	4.2	240	1	ECX1_METTH	O26779 methanobact
25	148.5	4.2	239	1	ECX1_METKA	Q8tyc1 methanopyru
26	147.5	4.1	240	1	RR41_CAEEL	Q17533 caenorhabdi
27	146	4.1	248	1	RNPH_LISIN	Q92ch1 listeria in
28	143.5	4.0	259	1	RNPH_MYCTU	Q10628 mycobacteri
29	143	4.0	245	1	RNPH_CORGL	Q8nmr4 corynebacte
30	142.5	4.0	544	1	THS_METTL	O93624 methanococc
31	140.5	3.9	247	1	RNPH_ANASP	Q8z0m2 anabaena sp
32	140	3.9	248	1	ECX1_THEAC	Q9hip2 thermoplasm
33	139.5	3.9	245	1	RR41_MOUSE	Q921i9 mus musculu
34	138	3.9	245	1	RNPH_STRCO	Q9s2h7 streptomyce
35	137.5	3.9	238	1	RNPH_YERPE	Q8zjp8 yersinia pe
36	137	3.9	248	1	ECX1_THEVO	Q97bz5 thermoplasm
37	137	3.9	276	1	ECX2_AERPE	Q9yc05 aeropyrum p
38	136.5	3.8	1575	1	IQG2_HUMAN	Q13576 homo sapien
39	136	3.8	238	1	RNPH_PASHA	Q9aes0 pasteurella
40	134	3.8	245	1	RR41_HUMAN	Q9npd3 homo sapien
41	132	3.7	238	1	RNPH_SALTY	P26155 salmonella
42	131	3.7	259	1	RNPH_MYCLE	P37939 mycobacteri
43	130	3.7	238	1	RNPH_ECOLI	P03842 escherichia
44	129.5	3.6	239	1	RNPH_PSEAE	P50597 pseudomonas
45	129	3.6	238	1	RNPH_HAEIN	P44444 haemophilus

ALIGNMENTS

RESULT 1
PNP_ECOLI
ID PNP_ECOLI STANDARD; PRT; 711 AA.
AC P05055; P78109;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
DE phosphorylase) (PNPase).
GN PNP OR B3164.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87083499; PubMed=2432069;
RA Regnier P., Grunberg-Manago M., Portier C.;
RT "Nucleotide sequence of the pnp gene of Escherichia coli encoding
RT polynucleotide phosphorylase. Homology of the primary structure of
RT the protein with the RNA-binding domain of ribosomal protein S1.";
RL J. Biol. Chem. 262:63-68(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE OF 1-196 FROM N.A.
RX MEDLINE=84297215; PubMed=6382163;
RA Portier C., Regnier P.;
RT "Expression of the rpsO and pnp genes: structural analysis of a DNA
RT fragment carrying their control regions.";
RL Nucleic Acids Res. 12:6091-6102(1984).
RN [4]
RP SEQUENCE OF 1-62 FROM N.A.
RX MEDLINE=86137413; PubMed=3005122;
RA Evans S., Dennis P.P.;
RT "Promoter activity and transcript mapping in the regulatory region
RT for genes encoding ribosomal protein S15 and polynucleotide
RT phosphorylase of Escherichia coli.";
RL Gene 40:15-22(1985).
RN [5]
RP SEQUENCE OF 1-12.
RC STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
RN [6]

```
RP   STRUCTURE BY NMR OF 617-692.
RX   MEDLINE=97160844; PubMed=9008164;
RA   Bycroft M., Hubbard T.J., Proctor M., Freund S.M., Murzin A.G.;
RT   "The solution structure of the S1 RNA binding domain: a member of an
RT   ancient nucleic acid-binding fold.";
RL   Cell 88:235-242(1997).
CC   -!- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC       polyribonucleotides processively in the 3' to 5' direction.
CC       Involved in the RNA degradosome, a multi-enzyme complex important
CC       in RNA processing and messenger RNA degradation.
CC   -!- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC       nucleoside diphosphate.
CC   -!- SUBUNIT: Homotrimer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- INDUCTION: In response to low temperature.
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   -!- SIMILARITY: Contains 1 S1 motif domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02638; AAA83905.1; -.
DR   EMBL; U18997; AAA57967.1; ALT_INIT.
DR   EMBL; AE000397; AAC76198.1; ALT_INIT.
DR   EMBL; X00761; CAA25332.1; -.
DR   EMBL; M14425; AAA24596.1; -.
DR   PIR; H65106; H65106.
DR   PDB; 1SRO; 01-APR-97.
DR   SWISS-2DPAGE; P05055; COLI.
DR   ECO2DBASE; C078.0; 6TH EDITION.
DR   EcoGene; EG10743; pnp.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000110; Ribosomal_S1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   PRINTS; PR00681; RIBOSOMALS1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Nucleotidyltransferase; RNA-binding; 3D-structure;
KW   Complete proteome.
FT   DOMAIN      557    591       KH.
FT   DOMAIN      622    690       S1 MOTIF.
FT   CONFLICT    357    357       G -> R (IN REF. 1).
FT   CONFLICT    450    450       L -> S (IN REF. 1).
FT   TURN        621    622
FT   STRAND      624    633
FT   TURN        634    635
FT   STRAND      636    640
FT   STRAND      648    648
FT   TURN        651    652
FT   HELIX       662    665
FT   TURN        668    669
FT   STRAND      671    678
FT   TURN        681    682
FT   STRAND      686    689
SQ   SEQUENCE   711 AA;  77101 MW;  785B7D54716FC2DE CRC64;

  Query Match             32.5%;  Score 1155;  DB 1;  Length 711;
  Best Local Similarity   39.9%;  Pred. No. 1e-62;
  Matches  269;  Conservative  118;  Mismatches  252;  Indels   36;  Gaps   11;
```

```
Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
              |   : : :| :|| |  : :|   |||| || | : |  | | | || |:|:::  |||
Db         12 GQHTVTLETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQDFFPLTVNYQERTYAAG 71

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
              ||| :: |||   |: | | :|:||| |||||| |:  : ||:  :::|:    ||::|:
Db         72 RIPGSFFRREGRPSEGETLIARLIDRPIRPLFPEGFVNEVQVIATVVSVNPQVNPDIVAM 131

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
              ||| ||||| ||:|||:|| |:| |: :||:|||: |:  | |:|||||  :: ::|:|
Db        132 IGASAALSLSGIPFNGPIGAARVGYINDQYVLNPTQDELKESKLDLVVAGT-EAAVLMVE 190

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSP---EIVKYT 270
              : |: : :    |:  | :  | :|| | :|||| |     |:  : | |    :
Db        191 SEAQLLSEDQMLGAVVFGHEQQQVVIQNINELVKEAG----KPRWDWQPEPVNEALNARV 246

Qy        271 HKLAMERL---YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVA 327
              ||  ||   | :    ||  |   |: |: :|  |  :    |  |:|: |  : :
Db        247 AALAEARLSDAYRI-----TDKQERYAQVDVIKSETIATLLAEDETLDENELGEILHAIE 301

Qy        328 KEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLES 387
              | | || ||    | |||:   :| :     :   ||||| ||:|| | | |  :
Db        302 KNVVRSRVLAGEPRIDGREKDMIRGLDVRTGVLPRTHGSALFTRGETQALVTATLGT--- 358

Qy        388 GIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP- 446
                :  ||:  : | :   |: || ||||:  | | |    |||:||| ||:: :  |:|
Db        359 -ARDAQVLDELMGERTDTFLFHYNFPPYSVGETGMVGSPKRREIGHGRLAKRGVLAVMPD 417

Qy        447 -RDFPPTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKG 505
               ||:|:|| ||: |||||||||| || ||||||:|||| :||||:|:||| : |
Db        418 MDKFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVKEGD---- 473

Qy        506 EIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAK 565
                :| :|:|||| ||: |||||:||:  ||:||| ||| :|  :|  |: ||  |:
Db        474 ---NYVVLSDILGDEDHLGDMDFKVAGSRDGISALQMDIKIEGITKEIMQVALNQAKGAR 530

Qy        566 KEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQV 625
               || :| : |: ||    |  | : |:::   |    :| ||  :: |  ||| ||
Db        531 LHILGVMEQAINAPRGDISEFAPRIHTIKINPDKIKDVIGKGGSVIRALTEETGTTIEIE 590

Qy        626 DEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNM 685
              |: |  : |       |  | ||       ::| | |||  :| | | |  | :
Db        591 DDGTVKIAATDGEKAKHAIRRIEEI-----TAEIEVGRVYTGKVTRIVDFGAFVAIGGGK 645

Qy        686 TAVLLHNTQLDNERL 700
               |:| :|: ::|:
Db        646 EG-LVHISQIADKRV 659
```

```
RESULT 2
PNP_BUCAP
ID   PNP_BUCAP      STANDARD;      PRT;   707 AA.
AC   Q8K9H5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
DE   phosphorylase) (PNPase).
GN   PNP OR BUSG361.
OS   Buchnera aphidicola (subsp. Schizaphis graminum).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Buchnera.
OX   NCBI_TaxID=98794;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22084549; PubMed=12089438;
RA   Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA   Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT   "50 million years of genomic stasis in endosymbiotic bacteria.";
RL   Science 296:2376-2379(2002).
CC   -!- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC       polyribonucleotides processively in the 3' to 5' direction.
```

```
CC       Involved in the RNA degradosome, a multi-enzyme complex important
CC       in RNA processing and messenger RNA degradation (By similarity).
CC   -!- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC       nucleoside diphosphate.
CC   -!- SUBUNIT: Homotrimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   -!- SIMILARITY: Contains 1 S1 motif domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE014112; AAM67914.1; -.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000110; Ribosomal_S1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   PRINTS; PR00681; RIBOSOMALS1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Nucleotidyltransferase; RNA-binding; Complete proteome.
FT   DOMAIN      553    612       KH.
FT   DOMAIN      622    690       S1 MOTIF.
SQ   SEQUENCE   707 AA;  78389 MW;  9D5367E96E91520A CRC64;

  Query Match           32.4%;  Score 1152;  DB 1;  Length 707;
  Best Local Similarity  38.5%;  Pred. No. 1.5e-62;
  Matches  264;  Conservative  128;  Mismatches  237;  Indels   56;  Gaps   14;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAV-SKTKPSPSQFMPLVVDYRQKAAAAG 93
              |   : : :| :|| |  : :    |||| || |  ||   |  :| || |:|::  | |
Db         12 GQHTITLETGIMARQATAAVMASMDDTAVFVTVVGEKTTNSSQKFFPLTVNYQERTYAVG 71

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
              |||  : |||   |: |||:|:||| |||||| |:  : |::  :::|:    ||:::|
Db         72 RIPGGFFRREGRPSENEILTARLIDRPIRPLFPKGFCNEIQIIATVVSVNPQINPDIISI 131

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
              ||| ||||| ||: ||:|| |:| :: :||:|||  :| || |:|||:|  :: ::|:|
Db        132 IGASAALSLSGIPFYGPIGAARVGFVNNQYVLNPTIDDMKSSFLDLVVSGT-QNAVLMVE 190

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTH-- 271
              | :: : :     ||  | :  | :|  |: |  |          ||    | |: |
Db        191 AESKVLSEDKILGAIMFGHQQQQVVINNIRSLSNEA-------SKL----PWIISYPEIN 239

Qy        272 --------KLA---MERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPE---ADPY 317
                     |||   :   | :|         :: |   |:    || :|  | |    |
Db        240 TELELKITKLAEKDISNAYLIF--------NKQERYEKLNFLKEEIIKLFFSENSNIDIS 291

Qy        318 EIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVL 377
              || : |  : | : |  :||   | |||:   :| :     :    |||:|| ||:|| |
Db        292 EIEDIFEKIEKNIVRKRILNNENRIDGREKDMIRALDIRTGVLPRTHGSSLFTRGETQSL 351

Qy        378 CTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALA 437
              :|| :       |:::    | |  ||: || ||||:  ||| |     |||:||| ||
Db        352 VSVTLGTSRDAQNLDELL----GDKTDNFLFHYNPPPYSVGEIGIVGSPKRREIGHGRLA 407

Qy        438 EKALYPVIPR--DFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIG 495
              :::|  |:|:  |||:|||: ||: |||||||||| || ||||||:|||| |||||:::|
Db        408 KRSLLAVMPKLDDFPYTIRIVSEITESNGSSSMASVCGASLALMDAGVPIKSAVAGISMG 467

Qy        496 LVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVM 555
              || : |        | ||:|||| ||: ||||||::|| :|||||| |||: ||  :|:
Db        468 LVKEGD-------KYVLLSDILGDEDHLGDMDFKVSGTEEGITALQMDIKIEGITNEIMR 520

Qy        556 EAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQ 615
               |: :|  |:  || :| : :||||    |  | :  :::   |    :| ||  :: |
Db        521 IALNKAKSARLHILNVMKQALSKPRNEISEFAPRIHKIKINPEKIKDVIGKGGSVIRMLT 580

Qy        616 AETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDT 675
               ||| |   |: |  : |   ::  :|:: |  |  ::   ::| | :|:  :| | |
Db        581 EETGTIIEIEDDGTIKISA---TIGEKAKNAIRRI--EEITAEIEVGRIYSGKVTRIVDF 635

Qy        676 GVMVKLYPNMTAVLLHNTQLDNERL 700
              |  : :       |:| :|: |:|:
Db        636 GAFISIGIGKEG-LVHISQISNKRV 659


RESULT 3
PNP_RICPR
ID   PNP_RICPR      STANDARD;      PRT;   745 AA.
AC   Q9ZD43;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
DE   phosphorylase) (PNPase).
GN   PNP OR RP504.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   -!- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC       polyribonucleotides processively in the 3' to 5' direction.
CC       Involved in the RNA degradosome, a multi-enzyme complex important
CC       in RNA processing and messenger RNA degradation (By similarity).
CC   -!- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC       nucleoside diphosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   -!- SIMILARITY: Contains 1 S1 motif domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235272; CAA14956.1; -.
DR   PIR; B71654; B71654.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
```

```
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Nucleotidyltransferase; RNA-binding; Complete proteome.
FT   DOMAIN       558    592       KH.
FT   DOMAIN       623    691       S1 MOTIF.
SQ   SEQUENCE   745 AA;  82339 MW;  885965045E76D2F2 CRC64;

  Query Match             32.1%;  Score 1140.5;  DB 1;  Length 745;
  Best Local Similarity   37.8%;  Pred. No. 8.2e-62;
  Matches  255;  Conservative  143;  Mismatches  236;  Indels   41;  Gaps   12;

Qy         39 LEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPS-QFMPLVVDYRQKAAAAGRIPT 97
              ||||:||:|| |: :  |: |:: :: | |    |       | || ::||: | :||:||
Db         16 LEISTGKIARQANAAVTVKMGNSILLCTCVVANKVKDGIGFFPLTINYREMAYSAGKIPG 75

Qy         98 NYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGAS 157
               : :||   ||:||| ||:||| |||||   : ::| | |::|: |     |:||| |||
Db         76 GFFKREGKASDREILVSRLIDRPIRPLFHQAFMHETHVTCSVLSYDPATPVDILAIIGAS 135

Qy        158 VALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAE 217
              |||:|  |:   | | ::|:|:||:|::|| : : :| |:|||||   | ::|:|: |
Db        136 AALSISPAPYLEIVAASKVGLINGEFVLNPTLELLKTSQLDLVVAGTEDS-VMMVESEAH 194

Qy        218 NILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMER 277
               : :     |:| | :  | :|: |::| ||     |  | |: ||      : |  :|:
Db        195 LLSEDKMLEAVKFGFESFQTVIKLIKELAKEAKKPKFEMQDLY-PS------SLKKEIEK 247

Qy        278 LYAVFTDYEHDKVSRDEAVNKIRLDTEEQL--------KEKFPEADPYEIIESFNVVAKE 329
              |:    :   :  |: |    : |  |: |          :|:   : |:|  :   :: :
Db        248 LFTKEVEQAFEIKSKQERSTDLALIYEKVLTHFVRDIENKKY---NNYQIESALKAISAD 304

Qy        330 VFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTF-DSLESG 388
              : |: :| :  | |||  | :| ::||| :  : ||||||  ||:|| | : ||  ||:
Db        305 ILRNKILEKNIRIDGRSTTDIRQIACEVGLLPSAHGSALFTRGETQSLVSTTFGTSLD-- 362

Qy        389 IKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIPR- 447
                :|:: :: |   : |||:| ||||: ||   :   :|||:||| || :|: |::|
Db        363 ---EQIVDSLEGEYKERFMLNYIFPPYSVNEAMPMKAPSRREVGHGKLAWRAINPILPNK 419

Qy        448 -DFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGE 506
               ||::||| :|  |||||||||: || ||||| :|||| : |||:|:|||         |
Db        420 VQFPYSIRVVAETTESNGSSSMATVCGSSLALMHAGVPIKAPVAGIAMGLVK-------E 472

Qy        507 IEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKK 566
               : :|:||:| ||| |||||:|||: ||||||  |||:  ||   ||:   |::|| : :
Db        473 SNKFAVLSDIIGDEDYFGDMDFKVAGTSSGITALQMDIKISGIDFKIIQIALEQARLGRL 532

Qy        567 EILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVD 626
               ||: ||| |||| :   :| |   ||::   |    :|||  :|::   :   |   |
Db        533 HILEQMNKVISKPNSELSKNAPSSTTVKIDKDKIKDIIGPGGKIIKEICETSNAKIDISD 592

Qy        627 EETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMT 686
              : | |::|     :  | | |  | :      | | ::  |: :: |:|  :    |
Db        593 DGTVSIYASDRDKIKIALDKIKAIAVEP-----EIGEIFNGTVMKVLDSGAFINYLGNKD 647

Qy        687 AVLLHNTQLDNERLN 701
              :| ::: : |::
Db        648 G-FVHISEISDARID 661


RESULT 4
PNP_YEREN
ID   PNP_YEREN      STANDARD;      PRT;   706 AA.
AC   O34275;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
DE   phosphorylase) (PNPase).
GN   PNP.
OS   Yersinia enterocolitica.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=630;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98294047; PubMed=9632258;
RA   Goverde R.L.J., Huis in't Veld J.H.J., Kusters H.G., Mooi F.R.;
RT   "The psychrotrophic bacterium Yersinia enterocolitica requires
RT   expression of pnp, the gene for polynucleotide phosphorylase, for
RT   growth at low temperature (5 degrees C).";
RL   Mol. Microbiol. 28:555-569(1998).
CC   -!- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC       polyribonucleotides processively in the 3' to 5' direction.
CC       Involved in the RNA degradosome, a multi-enzyme complex important
CC       in RNA processing and messenger RNA degradation (By similarity).
CC   -!- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC       nucleoside diphosphate.
CC   -!- SUBUNIT: Homotrimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   -!- SIMILARITY: Contains 1 S1 motif domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y10692; CAA71697.1; ALT_INIT.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Nucleotidyltransferase; RNA-binding.
FT   DOMAIN       557    591       KH.
FT   DOMAIN       622    690       S1 MOTIF.
SQ   SEQUENCE   706 AA;  76284 MW;  9CD82CD348C3AD4A CRC64;

  Query Match             31.9%;  Score 1136;  DB 1;  Length 706;
  Best Local Similarity   39.8%;  Pred. No. 1.4e-61;
  Matches  268;  Conservative  115;  Mismatches  259;  Indels   32;  Gaps   11;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
              |   : | :| :|| |  : :|   |||| || | : |  | | | || |:|:::  |||
Db         12 GQHTVTIETGMMARQATAAVMVSMDDTAVFVTVVGQKKAKPGQSFFPLTVNYQERTYAAG 71

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
              ||| :: |||   |: | ||||:||| ||||||  :  : ||:  :::|:    ||::|:
Db         72 RIPGSFFRREGRPSEGETLTSRLIDRPIRPLFPDSFLNEVQVIATVVSVNPQINPDIVAL 131

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
              ||| ||||| ||:|||:|| |:| |: :||:|||   |:  | |:|||||     ::|:|
Db        132 IGASAALSLSGIPFNGPIGAARVGFINDQYVLNPTTDELKESRLDLVVAGT-AGAVLMVE 190

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTH-- 271
              : |: : :     |:  | :  | :|: |  || | |     |:  :   | : :  |
Db        191 SEADILSEDQMLGAVVFGHEQQQVVIENINALVAEAG----KPKWDWHAEP-VNEALHAR 245

Qy        272 --KLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAKE 329
                :||  ||   :   |  | |   |: |: |  | |  :    |  || :    | |:
Db        246 VAELAAARLGDAYRITE--KQERYTQVDAIKADVTEALLAQDDTLDAAEIQDILGSVEKD 303
```

```
Qy        330 VFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGI 389
              | || ||    | |||:   :| :      :    |||||| ||:|| | | |  :
Db        304 VVRSRVLRGEPRIDGREKDMIRGLDVRTGVLPRTHGSALFTRGETQALVTATLGTARDAQ 363

Qy        390 KSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVI--PR 447
               |:::    | :  :|:||| ||||:  | | |    |||:||| ||:: :  |:  |
Db        364 NIDELM----GERTDSFLLHYNFPPYSVGETGMVGSPKRREIGHGRLAKRGVLAVMPSPS 419

Qy        448 DFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEI 507
              :||:|:|| ||: |||||||||| || ||||||:|||| :||||:|:|||        |
Db        420 EFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVK-------ED 472

Qy        508 EDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKE 567
              |:: :|:|||| ||: |||||:||:  |||||| |||: ||  :|:  |: ||   |:
Db        473 ENFVVLSDILGDEDHLGDMDFKVAGSRDGITALQMDIKIEGITREIMQVALNQAKGARLH 532

Qy        568 ILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDE 627
              || :| : || ||    |  | : |:::   |    :| ||  :: |  ||| ||    :
Db        533 ILGVMEQAISTPRGDISEFAPRIYTMKINPEKIKDVIGKGGSVIRALTDETGTTIDIEAD 592

Qy        628 ETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMTA 687
              | : |  .  |   | ||      ::|   :|   :| | | |  | :
Db        593 GTIKIAATDGDKAKHAIRRIEEI-----TAEIEVNRIYAGKVTRIVDFGAFVAIGGGKEG 647

Qy        688 VLLHNTQLDNERLN 701
               |:| :|: ::|::
Db        648 -LVHISQIADKRVD 660


RESULT 5
PNP_HAEIN
ID   PNP_HAEIN      STANDARD;      PRT;   709 AA.
AC   P44584;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
DE   phosphorylase) (PNPase).
GN   PNP OR HI0229.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC       polyribonucleotides processively in the 3' to 5' direction.
CC       Involved in the RNA degradosome, a multi-enzyme complex important
CC       in RNA processing and messenger RNA degradation (By similarity).
CC   -!- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC       nucleoside diphosphate.
CC   -!- SUBUNIT: Homotrimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   -!- SIMILARITY: Contains 1 S1 motif domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32709; AAC21898.1; -.
DR   PIR; E64056; E64056.
DR   HSSP; P05055; 1SRO.
DR   TIGR; HI0229; -.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Nucleotidyltransferase; RNA-binding; Complete proteome.
FT   DOMAIN      552    622       KH.
FT   DOMAIN      621    689       S1 MOTIF.
SQ   SEQUENCE   709 AA;  77006 MW;  BD9E08EA6236C860 CRC64;

  Query Match            31.9%;  Score 1134;  DB 1;  Length 709;
  Best Local Similarity  36.8%;  Pred. No. 1.9e-61;
  Matches  249;  Conservative  137;  Mismatches  251;  Indels   40;  Gaps   12;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
              |   : : :| :|| |  : :    || | || |:|      | | || |:|:::  |||
Db         11 GQHTVTLETGAIARQATAAVMASMDDTTVFVTVVAKKDVKEGQDFFPLTVNYQERTYAAG 70

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
              :||  : :||   |: | | :|:||| |||||| |:| : ||:  :::|:    ||::|:
Db         71 KIPGGFFKREGRPSEGETLIARLIDRPIRPLFPEGFFNEIQVVATVVSVNPQISPDLVAM 130

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
               ||| ||:|| :|:|||:|| |:| || ::|:|||  |    | |:||||  |: ::|:|
Db        131 IGASAALTLSGVPFNGPIGAARVGFIDNQFVLNPTMAEQKQSRLDLVVAGTDKA-VLMVE 189

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTK---RTPQKLFTPSPEIVKYT 270
              : |: : ::    |:  | :  | ::: |::  || |  :     ||   |: :::
Db        190 SEADILTEEQMLAAVVFGHQQQQVVVEAIKEFAKEAGKPRWDWVAPQ----PNTDLINKV 245

Qy        271 HKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAKEV 330
              :|  ||   :  |  |  | | :: |: |   |:  : |    :|:: |  :  ::
Db        246 KAIAEARLGDAYRITE--KQLRYEQIDAIKADVIAQITAEDEEXSEGKIVDIFTALESQI 303

Qy        331 FRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIK 390
              |  ::    | ||| : ::| :      :    ||||:| ||:|| |    |      | :
Db        304 VRGRIIAGEPRIDGRTVDTVRALDICTGVLPRTHGSAIFTRGETQALAVATL-----GTE 358

Qy        391 SD-QVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP--R 447
              | |:|  : | :  :|: || ||||:  | | :    |||:||| ||:: :  |:|
Db        359 RDAQIIDELTGERQDHFLFHYNFPPYSVGETGMIGSPKRREIGHGRLAKRGVAAVMPSLA 418

Qy        448 DFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEI 507
              :||: :|| ||: |||||||||| || ||||||:|||| :||||:|:|||        |
Db        419 EFPYVVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVK-------EE 471

Qy        508 EDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKE 567
              | : :|:|||| ||: |||||:||| :|:|||| ||:  ||   :|:  |: ||   |:
Db        472 EKFVVLSDILGDEDHLGDMDFKVAGTREGVTALQMDIKIEGITPEIMQIALNQAKSARMH 531

Qy        568 ILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDE 627
              || :| : |  |||   :  | : |:::   |   :| ||  :: |  ||| :|   |:
Db        532 ILGVMEQAIPAPRADISDYAPRIYTMKIDPKKIKDVIGKGGATIRSLTEETGTSIDIDDD 591

Qy        628 ETFSVFA----PTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYP 683
              |  : |       :||    : : |:          | | :|   :| : | |  | :
```

```
Db        592 GTVKIAAVDSNAAKNVMGRIEEIVAEV---------EAGVIYKGKVTRLADFGAFVAIVG 642

Qy        684 NMTAVLLHNTQLDNERL 700
              |    |:| :|:  ||:
Db        643 NKEG-LVHISQIAEERV 658


RESULT 6
PNP_BACSU
ID   PNP_BACSU      STANDARD;      PRT;   704 AA.
AC   P50849;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
DE   phosphorylase) (PNPase) (Vegetative protein 15) (VEG15).
GN   PNPA OR COMR.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96423178; PubMed=8825779;
RA   Luttinger A., Hahn J., Dubnau D.;
RT   "Polynucleotide phosphorylase is necessary for competence development
RT   in Bacillus subtilis.";
RL   Mol. Microbiol. 19:343-356(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE OF 1-8 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97261879; PubMed=9108293;
RA   Coquard D., Huecas M., Ott M., van Dijl J.M., van Loon A.P.,
RA   Hohmann H.P.;
RT   "Molecular cloning and characterisation of the ribC gene from
RT   Bacillus subtilis: a point mutation in ribC results in riboflavin
RT   overproduction.";
RL   Mol. Gen. Genet. 254:81-84(1997).
RN   [4]
RP   SEQUENCE OF 1-20.
RC   STRAIN=168 / IS58;
RX   MEDLINE=97443988; PubMed=9298659;
RA   Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA   Hecker M.;
RT   "First steps from a two-dimensional protein index towards a response-
RT   regulation map for Bacillus subtilis.";
RL   Electrophoresis 18:1451-1463(1997).
CC   -!- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC       polyribonucleotides processively in the 3' to 5' direction.
CC       Necessary for competence development in Bacillus subtilis. May be
CC       Necessary for modification of the srfA transcript (stabilization
CC       or translation activation).
CC   -!- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC       nucleoside diphosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   -!- SIMILARITY: Contains 1 S1 motif domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U29668; AAC43595.1; -.
DR   EMBL; Z99112; CAB13542.1; -.
DR   EMBL; Z80835; CAB02561.1; -.
DR   PIR; S70691; S70691.
DR   HSSP; P05055; 1SRO.
DR   SubtiList; BG11491; pnpA.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Competence; Transferase; Nucleotidyltransferase; RNA-binding;
KW   Complete proteome.
FT   INIT_MET      0      0
FT   DOMAIN      553    612       KH.
FT   DOMAIN      622    690       S1 MOTIF.
SQ   SEQUENCE   704 AA;  77332 MW;  0E305B6B9B0A7B07 CRC64;

  Query Match             31.8%;  Score 1132;  DB 1;  Length 704;
  Best Local Similarity   37.4%;  Pred. No. 2.5e-61;
  Matches  251;  Conservative  144;  Mismatches  253;  Indels   24;  Gaps   10;

Qy         32 VDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVVDYRQKAAA 91
              :|    | | : :|:||: |:|: ::: |||||: || :  :| |  | || |:| ::  |
Db          9 IDWAGRTLTVETGQLAKQANGAVMIRYGDTAVLSTATASKEPKPLDFFPLTVNYEERLYA 68

Qy         92 AGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVL 151
                |:||  :::||    |:| :| ||:||| |||||  |:  : ||:  :::||     ::
Db         69 VGKIPGGFIKREGRPSEKAVLASRLIDRPIRPLFADGFRNEVQVISIVMSVDQNCSSEMA 128

Qy        152 AINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVM 211
              |: |:|:|||:||||: ||:  | :| || ::::|||  ::  | :||||||  |  | |
Db        129 AMFGSSLALSVSDIPFEGPIAGVTVGRIDDQFIINPTVDQLEKSDINLVVAGT-KDAINM 187

Qy        212 LEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTH 271
              :|| |: : ::    ||  | :  :::|   :::|   |  :::  |||    |: :
Db        188 VEAGADEVPEEIMLEAIMFGHEEIKRLIAFQEEIVAAVG-KEKSEIKLFEIDEELNEKVK 246
```

```
Qy        272 KLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPY-EIIESFNVVAKEV 330
Db        247 ALAEEDLLKAI--QVHEKHAREDAINEVKNAVVAKFEDEEHDEDTIKQVKQILSKLVKNE 304

Qy        331 FRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIK 390
Db        305 VRRLITEEKVRPDGRGVDQIRPLSSEVGLLPRTHGSGLFTRGQTQALSVCTLGAL----G 360

Qy        391 SDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP--RD 448
Db        361 DVQILDGLGVEESKRFMHHYNFPQFSVGETGPMRGPGRREIGHGALGERALEPVIPSEKD 420

Qy        449 FPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEIE 508
Db        421 FPYTVRLVSEVLESNGSTSQASICASTLAMMDAGVPIKAPVAGIAMGLV-----KSG--E 473

Qy        509 DYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKEI 568
Db        474 HYTVLTDIQGMEDALGDMDFKVAGTEKGVTALQMDIKIEGLSREILEEALQQAKKGRMEI 533

Qy        569 LQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDEE 628
Db        534 LNSMLATLSESRKELSRYAPKILTMTINPDKIRDVIGPSGKQINKIIEETGVKIDIEQDG 593

Qy        629 TFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMTAV 688
Db        594 TIFISSTDESGNQKAKKII-----EDLVREVEVGQLYLGKVKRIEKFGAFVEIFSGKDG- 647

Qy        689 LLHNTQLDNERL 700
Db        648 LVHISELALERV 659
```

```
RESULT 7
PNP_BUCAI
ID   PNP_BUCAI      STANDARD;      PRT;   707 AA.
AC   P57454;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
DE   phosphorylase) (PNPase).
GN   PNP OR BU373.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tokyo 1998;
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC       polyribonucleotides processively in the 3' to 5' direction.
CC       Involved in the RNA degradosome, a multi-enzyme complex important
CC       in RNA processing and messenger RNA degradation (By similarity).
CC   -!- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC       nucleoside diphosphate.
CC   -!- SUBUNIT: Homotrimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   -!- SIMILARITY: Contains 1 S1 motif domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP001119; BAB13077.1; -.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Nucleotidyltransferase; RNA-binding; Complete proteome.
FT   DOMAIN      557    591       KH.
FT   DOMAIN      622    690       S1 MOTIF.
SQ   SEQUENCE   707 AA;  78234 MW;  75FA1BBE08105B7A CRC64;
```

```
  Query Match           31.3%;  Score 1115;  DB 1;  Length 707;
  Best Local Similarity  37.2%;  Pred. No. 2.7e-60;
  Matches  254;  Conservative  142;  Mismatches  237;  Indels   50;  Gaps   13;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
Db         12 GQHTITLETGVIARQANAAVMASMDETAVFVTVVGQKKIHTGQKFFPLTVNYQERTYAAG 71

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
Db         72 RIPGGFFRREGRPSENEILTARLIDRPLRPLFPKKFLNEIQIIATVVSVNPQINPDIISI 131

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
Db        132 IGASAALSLSGIPFYGPVGAARVGYINNQYILNPISDDMKNSSLDLVVSGT-QNAILMVE 190

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKY--TH 271
Db        191 AESKILSEEKILGAIIFGHQQQQVVINNIRSLSNEA-------SKL----PWVISYPETN 239

Qy        272 KLAMERL-----------YAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEII 320
Db        240 KTLELKIINSFEKNISDAYVIF-----NKQDRIEKLNSIKENIIKLFLDENSNIDTLEIE 294

Qy        321 ESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTV 380
Db        295 DIFQKIEKKVVRKRILSNQTRIDGREKDMIRALDVRTGILPRTHGSALFTRGETQSLVSV 354

Qy        381 TFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKA 440
Db        355 TLGTSRDAQNLDELL----GDRIDNFLFHYNFPPYSVGEIGMVGSPKRREIGHGRLAKRS 410

Qy        441 LYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVT 498
Db        411 LLAVMPTLENFPYTIRVVSEITESNGSSSMASVCGASLALMDAGVPIKSAVAGISMGLVK 470

Qy        499 KTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAI 558
Db        471 -------EGNQHVLLSDILGDEDHLGDMDFKVAGTEEGITALQMDMKIEGITNEIIHSAL 523

Qy        559 QQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAET 618
Db        524 NEARLARLHILNVMNQALNESRSEISEFAPRIHIIKINPEKIKDVIGKGGSVIRMLTEET 583

Qy        619 GVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVM 678
Db        584 GTIIEIEDDGTVKI---SSTVKEKAKNAIRRI--KEITAEIEVGRIYSGKVTRIVDFGAF 638

Qy        679 VKLYPNMTAVLLHNTQLDNERLN 701
Db        639 VSIGLGKEG-LVHISQISDKRVD 660
```

```
RESULT 8
PNP_PHOLU
ID   PNP_PHOLU      STANDARD;      PRT;   709 AA.
AC   P41121;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
DE   phosphorylase) (PNPase) (CAP87K).
GN   PNP OR PPH.
OS   Photorhabdus luminescens (Xenorhabdus luminescens).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Photorhabdus.
OX   NCBI_TaxID=29488;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-13.
RC   STRAIN=K122;
RX   MEDLINE=94266731; PubMed=8206856;
RA   Clarke D.J., Dowds B.C.A.;
RT   "The gene coding for polynucleotide phosphorylase in Photorhabdus sp.
RT   strain K122 is induced at low temperatures.";
RL   J. Bacteriol. 176:3775-3784(1994).
CC   -!- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC       polyribonucleotides processively in the 3' to 5' direction.
CC   -!- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC       nucleoside diphosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- INDUCTION: In response to low temperature.
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   -!- SIMILARITY: Contains 1 S1 motif domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X76069; CAA53671.1; -.
DR   PIR; S38883; S38883.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000110; Ribosomal_S1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   PRINTS; PR00681; RIBOSOMALS1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Nucleotidyltransferase; RNA-binding.
FT   DOMAIN       553     612       KH.
FT   DOMAIN       622     690       S1 MOTIF.
SQ   SEQUENCE   709 AA;  76745 MW;  EBFCC559FC921E6B CRC64;

  Query Match             31.1%;  Score 1106;  DB 1;  Length 709;
  Best Local Similarity   38.3%;  Pred. No. 9.5e-60;
  Matches  258;  Conservative  126;  Mismatches  256;  Indels   34;  Gaps   12;

Qy         35 GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAG 93
              |    : | :| :|| |  : :|   |||| || | : |    | | || |:|::: |||
Db         12 GQHTVTIETGMMARQATAAVMVNMDDTAVFVTVVGQKKVKAGQDFFPLTVNYQERTYAAG 71

Qy         94 RIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAI 153
```

```
              ||| :: |||    : | | :|:||| :||||| |:  : ::: :::|:    ||::|:
Db         72 RIPGSFFRREGRPGEGETLVARLIDRPLRPLFPEGFLNEVRIVATVVSVNPQINPDIVAM 131

Qy        154 NGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLE 213
              ||| ||:|| ||:|||:|| |:| |: :||:||| |: :| |:|||:|     ::|:|
Db        132 IGASAALALSGIPFNGPIGAARVGYINDQYVLNPTSDELKNSRLDLVVSGT-AGAVLMVE 190

Qy        214 ASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTH-- 271
              : |: : ::    |:  |    | :|  |  | | | |       : | | : :  |
Db        191 SEADLLTEEQMLGAVVFGHDQQQVVIDNINALAAEAGKEKWD----WVPEP-VNQALHDR 245

Qy        272 --KLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAKE 329
                :||  ||   :   |  |  |  |: |: :   | |:   :  || |    : |
Db        246 VAELAESRLGDAYRITE--KQERYAQVDAIKDEVTAALLEQDETLEEAEIHEILGSLEKN 303

Qy        330 VFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGI 389
              | || ||:   | |||:   :| :     :    ||||||  ||:|| | | |      |
Db        304 VVRSRVLSGEPRIDGREKDMVRALDVRTGVLPRTHGSALFTRGETQALVTATL-----GT 358

Qy        390 KSD-QVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIPR- 447
              : | |:|  : | :   |:||| ||||:  | | :    |||:||| ||:: :  |:|:
Db        359 ERDAQIIDELMGERTDRFLLHYNFPPYSVGETGMMGSPKRREIGHGRLAKRGVLAVMPKA 418

Qy        448 -DFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGE 506
               :||:|:|| ||: |||||||||| || ||||||:|||| :||||:|:||| : |
Db        419 NEFPYTVRVVSEITESNGSSSMASVCGASLALMDAGVPIKAAVAGIAMGLVKEGD----- 473

Qy        507 IEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKK 566
              :: :|:|||| ||: ||||||:||: :||:||| |||: ||  :|:  |: ||  |:
Db        474 --NFVVLSDILGDEDHLGDMDFKVAGSCEGISALQMDIKIEGITREIMQVALNQAKGARL 531

Qy        567 EILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVD 626
              || :| : |: ||    :  | : |:::   |    :| ||  :: |  ||| ||   |
Db        532 HILSVMEQAITTPRDDISQFAPRIHTIKINPDKIKDVIGKGGSVIRALTEETGTTIEIED 591

Qy        627 EETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMT 686
              : |  : |      |   | ||      ::| | :|   :| | | |  | :
Db        592 DGTVKIAATDGEKAKHAISRIEEI-----TAEIEVGRIYAGKVTRIVDFGAFVAIGGGKE 646

Qy        687 AVLLHNTQLDNERL 700
                |:| :|: ::|:
Db        647 G-LVHISQIADKRV 659
```

```
RESULT 9
PNP_PSEPU
ID   PNP_PSEPU      STANDARD;      PRT;   701 AA.
AC   O87792;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide
DE   phosphorylase) (PNPase).
GN   PNP.
OS   Pseudomonas putida.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TMB;
RA   Favaro R., Deho' G.;
RT   "Identification and cloning of genes involved in RNA turnover in
RT   Pseudomonas putida.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Involved in mRNA degradation. Hydrolyzes single-stranded
CC       polyribonucleotides processively in the 3' to 5' direction.
CC       Involved in the RNA degradosome, a multi-enzyme complex important
CC       in RNA processing and messenger RNA degradation (By similarity).
CC   -!- CATALYTIC ACTIVITY: {RNA}(N+1) + phosphate = {RNA}(N) + a
CC       nucleoside diphosphate.
CC   -!- SUBUNIT: Homotrimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
```

```
CC   -!- SIMILARITY: Contains 1 KH domain.
CC   -!- SIMILARITY: Contains 1 S1 motif domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y18132; CAA77048.1; -.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR000110; Ribosomal_S1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   PRINTS; PR00681; RIBOSOMALS1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Nucleotidyltransferase; RNA-binding.
FT   DOMAIN      558    592       KH.
FT   DOMAIN      623    691       S1 MOTIF.
SQ   SEQUENCE   701 AA;  74984 MW;  5DB8F1A446DF0F2C CRC64;

  Query Match           30.0%;  Score 1068;  DB 1;  Length 701;
  Best Local Similarity   36.3%;  Pred. No. 1.9e-57;
  Matches  247;  Conservative  130;  Mismatches  259;  Indels   44;  Gaps   14;

Qy         35 GNRKLEISSGKLARFADGSAVVQ-SGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAA 92
              |   : : :|::|| | |: :|     |  |:|| |   :  | : | || | |::|  ||
Db         11 GQSTVTLETGRIARQATGAVLVTVDNDVTVLVTVVGAKQADPGKGFFPLSVHYQEKTYAA 70

Qy         93 GRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLA 152
              |:||  : :||   |:|| ||||:||| |||||| |:  : ||:| :::     :||: |
Db         71 GKIPGGFFKREGRPSEKETLTSRLIDRPIRPLFPEGFMNEVQVVCTVVSTSKKTDPDIAA 130

Qy        153 INGASVALSLSDIPWNGPVGAVRIGIIDGE-YVVNPTRKEMSSSTLNLVVAGAPKSQIVM 211
              : | | ||::| ||: ||:|| |:   :   |::||| :::::|:|::|||     ::|
Db        131 MIGTSAALAISGIPFEGPIGAARVAFHESTGYLLNPTYEQLAASSLDMVVAGT-SDAVLM 189

Qy        212 LEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTH 271
              :|: |: : :     |:       | :|| :::|  | |     |   : |:    :  :
Db        190 VESEAQELTEDQMLGAVLFAHDEFQAVIQAVKELAAEAG----KPTWDWKPAVANTELFN 245

Qy        272 KLAMERLYAVFTDY----EHDKVS-----RDEAVNKIRLDTEEQLKEKFPEADPYEIIES 322
              :  |   ||   |    : |: :     ||:|| |         :|  | |   |: |
Db        246 AIRAEFGEAVSQGYTITVKADRYARLGELRDQAVAKF------SGEEGQPSAS--EVKEI 297

Qy        323 FNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTF 382
              |  :     |  ::|   | ||||  ::| :: || :    ||||||  ||:|| |   |
Db        298 FGEIEYRTVRENIVNGKPRIDGRDNKTVRPLNIEVGVLPKTHGSALFTRGETQALVVATL 357

Qy        383 DSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALY 442
              :     :  |::  : | |   ||||| |||::  | |:: |  |||:||| || :::
Db        358 GT----ARDAQLLDTLEGEKKDPFMLHYNFPPFSVGECGRMGGAGRREIGHGRLARRSVQ 413

Qy        443 PVIPRD--FPPTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKT 500
              ::|    ||:|||| ||: |||||||||| || ||||||:|||: : |||:|:||| :
Db        414 AMLPAADVFPYTIRVVSEITESNGSSSMASVCGASLALMDAGVPMKAPVAGIAMGLVKEG 473

Qy        501 DPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQ 560
              |       : :|||||| ||: ||||||:||| ||:|||| |||: ||  :|:  |: |
Db        474 D-------KFAVLTDILGDEDHLGDMDFKVAGTAKGVTALQMDIKINGITEEIMEIALGQ 526

Qy        561 ASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGV 620
              |  |: ||  ||: | : |     | | : :::   |    :| ||  :: :  ||
Db        527 ALEARLNILGQMNQVIGQSRTELSANAPTMIAMKIDTDKIRDVIGKGGATIRAICEETKA 586

Qy        621 TISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVK 680
              :|   |: :  :|  |       |:  |  |       : | | :|   :  | | |  |
Db        587 SIDIEDDGSIKIFGETKEAADAAKQRILGI-----TAEAEIGKIYVGKVERIVDFGAFVN 641

Qy        681 LYPNMTAVLLHNTQLDNERL 700
              : |     |:| : | : |:
Db        642 ILPGKDG-LVHISMLSDARV 660


RESULT 10
ECX1_SULSO
ID   ECX1_SULSO     STANDARD;      PRT;   248 AA.
AC   Q9UXC2;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Probable exosome complex exonuclease 1 (EC 3.1.13.-).
GN   SSO0735.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=20165948; PubMed=10701121;
RA   Charlebois R.L., Singh R.K., Chan-Weiher C.C.-Y., Allard G., Chow C.,
RA   Confalonieri F., Curtis B., Duguet M., Erauso G., Faguy D.,
RA   Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,
RA   Kushwaha N., Lafleur E., Medina N., Peng X., Penny S.L., She Q.,
RA   St Jean A., van der Oost J., Young F., Zivanovic Y., Doolittle W.F.,
RA   Ragan M.A., Sensen C.W.;
RT   "Gene content and organization of a 281-kbp contig from the genome of
RT   the extremely thermophilic archaeon, Sulfolobus solfataricus P2.";
RL   Genome 43:116-136(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=11427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC   -!- FUNCTION: Probably involved in the 3'->5' degradation of a variety
CC       of RNA species (Potential).
CC   -!- SUBUNIT: Component of the archaeal exosome multienzyme
CC       ribonuclease complex (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y18930; CAB57569.1; -.
DR   EMBL; AE006698; AAK41031.1; -.
DR   PIR; H90221; H90221.
DR   HAMAP; MF_00591; -; 1.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   Pfam; PF01138; RNase_PH; 1.
```

```
DR   Pfam; PF03725; RNase_PH_C; 1.
KW   Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ   SEQUENCE   248 AA;  27578 MW;  EAB2C2C89DD5854C CRC64;

  Query Match             5.9%;  Score 209;  DB 1;  Length 248;
  Best Local Similarity   27.2%;  Pred. No. 6.4e-06;
  Matches   72;  Conservative   45;  Mismatches   86;  Indels   62;  Gaps   10;

Qy        334 IVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQ 393
              ::|:: || |||    ||::  |: : |   |||:|: | |                :
Db         12 LILDDGKRTDGRKPDELRSIKIELGVLKNADGSAIFEMGNT-----------------K 53

Qy        394 VITAINGIKDKN-----------FMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKAL- 441
              | |: | |: :             : |   |::|:|  |    :|||:    :  :|| 
Db         54 AIAAVYGPKEMHPRHLSLPDRAVLRVRYHMTPPSTDE-RKNPAPSRREIELSKVIREALE 112

Qy        442 YPVIPRDFPFT-IRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKT 500
               |:   || | | | :|:|:::  | : |    |||| |:|:|:   :||||:|
Db        113 SAVLVELFPRTAIDVFTEILQADAGSRLVSLMAASLALADAGIPMRDLIAGVAVGKA--- 169

Qy        501 DPEKGEIEDYRLLTDILGIEDYNGDMDFKIA---GTNKGITALQA-------------DI 544
                  |  :: |:   ||  |: |  ||     |: :|  |                 |:
Db        170 --------DGVIILDLNETEDMWGEADMPIAMMPSLNQ-VTLFQLNGSMTPDEFRQAFDL 220

Qy        545 KLPGIPI--KIVMEAIQQASVAKKE 567
               : || |   :  ||::   |  ||
Db        221 AVKGINIIYNLEREALKSKYVEFKE 245


RESULT 11
ECX1_AERPE
ID   ECX1_AERPE     STANDARD;      PRT;   246 AA.
AC   Q9YC03;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Probable exosome complex exonuclease 1 (EC 3.1.13.-).
GN   APE1447.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC   Desulfurococcaceae; Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
CC   -!- FUNCTION: Probably involved in the 3'->5' degradation of a variety
CC       of RNA species (Potential).
CC   -!- SUBUNIT: Component of the archaeal exosome multienzyme
CC       ribonuclease complex (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP000061; BAA80445.1; -.
DR   PIR; G72623; G72623.
DR   HAMAP; MF_00591; -; 1.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   Pfam; PF01138; RNase_PH; 1.
DR   Pfam; PF03725; RNase_PH_C; 1.
KW   Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ   SEQUENCE   246 AA;  26670 MW;  9548CEA2BC2F4BB6 CRC64;

  Query Match             5.9%;  Score 208.5;  DB 1;  Length 246;
  Best Local Similarity   27.1%;  Pred. No. 6.8e-06;
  Matches   69;  Conservative   47;  Mismatches   90;  Indels   49;  Gaps    8;

Qy        335 VLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQV 394
              :: : :| |||     || |  :| :      |||| : |:|:||
Db         11 LIRDGRRHDGRLPEDLRPVRMQVGILHNADGSALVEFGRTRVL----------------- 53

Qy        395 ITAINGIKDKN-----------FMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYP 443
               |: | :: :             : |   |::|:|  |    :|||:    :  :|| |
Db         54 -AAVYGPREPHQRFYVLPDRAALRVRYHMAPFSTDE-RKSPAPSRREIELSKVVREALEP 111

Qy        444 VI-PRDFPFT-IRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTD 501
              |:   :|| | | |  |||:::| :  |:    |||| |:|:|: : | |||:| :
Db        112 VVLAEEFPRTVIDVFLEVLQADGGTRTAAVTAASLALADAGIPMRALVGGVAVGKIQGV- 170

Qy        502 PEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNK--GITALQADIKLPG----IPIKIVM 555
                    |: |:  :||  |: |  :|       || || :  | |      : : :
Db        171 ----------LVVDVDELEDMYGEADMPVAAAPDIGEITLLQLNGVLTGEEFRTALAMAL 220

Qy        556 EAIQQASVAKKEILQ 570
               || :    :|| ::
Db        221 RAIDRVVEMEKEAIR 235


RESULT 12
ECX1_METMA
ID   ECX1_METMA     STANDARD;      PRT;   493 AA.
AC   Q8PTT8;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Probable exosome complex exonuclease 1 (EC 3.1.13.-).
GN   MM2623.
OS   Methanosarcina mazei (Methanosarcina frisia).
OC   Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC   Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2209;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Goe1 / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX   MEDLINE=22120827; PubMed=12125824;
RA   Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA   Martinez-Arias R., Henne A., Wiezer A., Baeumer S., Jacobi C.,
RA   Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
RA   Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA   Fritz H.-J., Gottschalk G.;
RT   "The genome of Methanosarcina mazei: evidence for lateral gene
RT   transfer between Bacteria and Archaea.";
RL   J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC   -!- FUNCTION: Probably involved in the 3'->5' degradation of a variety
CC       of RNA species (Potential).
CC   -!- SUBUNIT: Component of the archaeal exosome multienzyme
CC       ribonuclease complex (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE013507; AAM32319.1; -.
DR   HAMAP; MF_00591; -; 1.
```

```
DR   InterPro; IPR001247; 3_ExoRNase.
DR   Pfam; PF01138; RNase_PH; 1.
DR   Pfam; PF03725; RNase_PH_C; 1.
KW   Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
FT   DOMAIN        1    254       PROBABLE EXOSOME COMPLEX EXONUCLEASE 1.
FT   DOMAIN      255    493       UNKNOWN.
SQ   SEQUENCE   493 AA;  55248 MW;  979A757BEF8DC090 CRC64;

  Query Match             5.4%;  Score 193.5;  DB 1;  Length 493;
  Best Local Similarity   22.5%;  Pred. No. 0.00015;
  Matches   78;  Conservative   59;  Mismatches  135;  Indels   75;  Gaps   11;

Qy        333 SIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSD 392
              ::: ::  | |||     :| :  || :      ||   : |: ::|  |         :|
Db          9 TLITDDGLRLDGRRADEIRPMKIEVGVLSRADGSCYLEWGRNKILVGVFGPREAHPRRSQ 68

Qy        393 QVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIPRD-FPF 451
              :  :|:           |    ::  :  :  | :|| :    :: :|  |||  : ||
Db         69 RADSAV-------IRYRYNMASFSVEDRAR-PGPSRRSIEISKVSREAFEPVIMAELFPK 120

Qy        452 T-IRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEIEDY 510
              | | :  |||:::  :  |:    |:|| |:|:|:   :   | | |           |
Db        121 TAIDIFVEVLQADAGTRTAAINASSIALADAGIPMKGLITSCAFGKV----------DG 169

Qy        511 RLLTDILGIEDYNGDMDFKIAGTNKG-ITALQADIKLPGIPIKIVMEAIQQASVAKKEIL 569
              ::: |:   ||  |: || :| |  | || :| |  |    ||  :| :::     ||||
Db        170 KIVLDLNKEEDNYGEADFPVAMTQDGEITLIQMDGNLTPDEIKQGLELVKKGC---KEIL 226

Qy        570 QIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDEET 629
              :|    :                  | ||  |                : :| |||
Db        227 EIQQAVL----------------------RKKFETP--------------VEEVSEET 248

Qy        630 P-------SVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATI 669
                      |  |:|         : |  :: :| ::|      | |
Db        249 APEKGAEKEVLEPSP-----VAAIVEETPEEAEEPEVEISEEVEAEI 290


RESULT 13
ECX1_SULTO
ID   ECX1_SULTO     STANDARD;      PRT;   247 AA.
AC   Q975G8;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Probable exosome complex exonuclease 1 (EC 3.1.13.-).
GN   ST0443.
OS   Sulfolobus tokodaii.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=111955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 10545 / 7;
RX   MEDLINE=21456156; PubMed=11572479;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA   Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA   Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA   Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA   Oshima T., Kikuchi H.;
RT   "Complete genome sequence of an aerobic thermoacidophilic
RT   Crenarchaeon, Sulfolobus tokodaii strain7.";
RL   DNA Res. 8:123-140(2001).
CC   -!- FUNCTION: Probably involved in the 3'->5' degradation of a variety
CC       of RNA species (Potential).
CC   -!- SUBUNIT: Component of the archaeal exosome multienzyme
CC       ribonuclease complex (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP000982; BAB65433.1; -.
DR   HAMAP; MF_00591; -; 1.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   Pfam; PF01138; RNase_PH; 1.
DR   Pfam; PF03725; RNase_PH_C; 1.
KW   Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ   SEQUENCE   247 AA;  27644 MW;  7EE042B7284C6599 CRC64;

  Query Match             5.4%;  Score 192;  DB 1;  Length 247;
  Best Local Similarity   25.5%;  Pred. No. 6.9e-05;
  Matches   68;  Conservative   48;  Mismatches  103;  Indels   48;  Gaps    8;

Qy        334 IVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQ 393
              ::|:   | |||    :| :  |: : |   |||:|: | |                 :
Db         13 LILDNGLRLDGRKPDEMRPIKIELGVLKNADGSAIFEMGNT-----------------K 54

Qy        394 VITAINGIKDKN----------FMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALY 442
              || |: | |: :              : |   |::|:|  |    :|||:    :  :||
Db         55 VIAAVYGPKEMHPRHLALPDRAVLRVRYHMTPFSTDE-RKNPAPSRREIELSKVIREALE 113

Qy        443 PVIPRD-FPFT-IRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKT 500
                | : || | | |  |||:::  : : |    |:|| |:|:|:   :||||:|
Db        114 STILVELFPRTVIDVFMEVLQADAGTRLVSLMAASMALADAGIPMRDLIAGVAVGKA--- 170

Qy        501 DPEKGEIEDYRLLTDILGIEDYNGDMDFKIA--GTNKGITALQADIKLPGIPIKIVMEAI 558
                    |  |: |:   ||  |: |  :|   : | :  || :  :     :   :|:
Db        171 --------DGVLVLDLNEPEDMWGEADMPVAMMPSLKQVALLQLNGNMTPQEFR---QAL 219

Qy        559 QQASVAKKEILQIMNKTISKPRASRKE 585
              : |    : |  :  : |     |  ||
Db        220 EMAQKGIETIYNLEKEAIRSKYAELKE 246


RESULT 14
ECX1_PYRAE
ID   ECX1_PYRAE     STANDARD;      PRT;   246 AA.
AC   Q8ZVM9;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Probable exosome complex exonuclease 1 (EC 3.1.13.-).
GN   PAE2207.
OS   Pyrobaculum aerophilum.
OC   Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC   Thermoproteaceae; Pyrobaculum.
OX   NCBI_TaxID=13773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX   MEDLINE=21664397; PubMed=11792869;
RA   Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA   Miller J.H.;
RT   "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT   aerophilum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC   -!- FUNCTION: Probably involved in the 3'->5' degradation of a variety
CC       of RNA species (Potential).
CC   -!- SUBUNIT: Component of the archaeal exosome multienzyme
CC       ribonuclease complex (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE009863; AAL64027.1; -.
DR   HAMAP; MF_00591; -; 1.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   Pfam; PF01138; RNase_PH; 1.
DR   Pfam; PF03725; RNase_PH_C; 1.
KW   Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ   SEQUENCE   246 AA;  27148 MW;  301145A699D3D62B CRC64;

  Query Match             5.2%;  Score 186;  DB 1;  Length 246;
  Best Local Similarity   27.6%;  Pred. No. 0.00016;
  Matches   61;  Conservative   32;  Mismatches   78;  Indels   50;  Gaps    7;

Qy        341 RCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAING 400
              | |||    :| |   | :     |||:  | |              : |: |
Db         14 RADGRTPDQMREVKIAVGVVSNADGSAMVSYGAT-----------------TAVAAVYG 55

Qy        401 ----------IKDKNFM-LHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVI---- 445
                       : |:  | : |   |::| :  |    :|||:    :  :|| | |
Db         56 PREMHPRHLSLPDRGVMRVRYHMAPFSTKDERKSPTPSRREIEISKILREALEPAIVLEQ 115

Qy        446 -PRDFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEK 504
               ||     | |  |:|:::||: :||    |||| |:|: :   | ||::|||
Db        116 YPRS---RIDVFVEILQADGSTRVASLTAASLALADAGIYMRDLVVGVSVGLV------- 165

Qy        505 GEIEDYRLLTDILGIEDYNGDMDFKIA--GTNKGITALQAD 543
                 |  :: |: |:||  |: |  :      |  | || |
Db        166 ----DGTVVLDLNGLEDQYGEGDLPLGYMPNLKRFTLLQLD 202


RESULT 15
ECX1_PYRAB
ID   ECX1_PYRAB     STANDARD;      PRT;   249 AA.
AC   Q9V119;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Probable exosome complex exonuclease 1 (EC 3.1.13.-).
GN   PYRAB06100 OR PAB0420.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GE5 / Orsay;
RX   PubMed=12622808;
RA   Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA   Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA   Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT   "An integrated analysis of the genome of the hyperthermophilic
RT   archaeon Pyrococcus abyssi.";
RL   Mol. Microbiol. 47:1495-1512(2003).
CC   -!- FUNCTION: Probably involved in the 3'->5' degradation of a variety
CC       of RNA species (Potential).
CC   -!- SUBUNIT: Component of the archaeal exosome multienzyme
CC       ribonuclease complex (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE RNASE PH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ248284; CAB49532.1; -.
DR   PIR; E75181; E75181.
DR   HAMAP; MF_00591; -; 1.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR002160; Kunitz_legume.
DR   Pfam; PF01138; RNase_PH; 1.
DR   Pfam; PF03725; RNase_PH_C; 1.
KW   Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ   SEQUENCE   249 AA;  27678 MW;  ACB0748E1BF074A6 CRC64;

  Query Match             4.9%;  Score 175.5;  DB 1;  Length 249;
  Best Local Similarity   27.5%;  Pred. No. 0.0007;
  Matches   72;  Conservative   40;  Mismatches   93;  Indels   57;  Gaps   11;

Qy        340 KRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAIN 399
              :| |||    || :  || : |  :|||  : |:                   :::| |:
Db         16 RRIDGRKKYELRPIKMEVGVLKNANGSAYIEWGK-----------------NKIIAAVY 57

Qy        400 GIKDKN-----------FMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYP-VIPR 447
              | :: :           : |   |::  |  |  | :|| :     : : || | :|
Db         58 GPRELHPKHLQRPDRAILRVRYNMAPFSVEE-RKKPGPDRRSIEISKVIKGALEPALILE 116

Qy        448 DFPPFT-IRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGE 506
              || | | |  |||:::  : :|     |||| |:|:|:   ||  | | :      :||
Db        117 MFPRTAIDVFIEVLQADAGTRVAGITAASLALADAGIPMRDLVAACAAGKI------EGE 170

Qy        507 IEDYRLLTDILGIEDYNGDMDFKIA--GTNKGITALQADIKLPGIPIK-IVMEAIQQASV 563
              |     : |:   ||  |: |  :|       || || |    |   |   :||:: |
Db        171 I-----VLDLNKEEDNYGEADVPVAIMPLKNDITLLQMD----GYLTKDEFIEAVKLAIK 221

Qy        564 AKKEILQIMNKTISKPRASRKE 585
                | : |       | | : ||
Db        222 GAKAVYQ-------KQREALKE 236


Search completed: January  8, 2004, 10:39:16
Job time : 18 secs
```

OM protein - protein search, using sw model

Run on: January 8, 2004, 10:31:51 ; Search time 42 Seconds
(without alignments)
4331.597 Million cell updates/sec

Title: US-09-907-907A-42
Perfect score: 3557
Sequence: 1 DGPFLLPRRDRALTQLQVRA.....TAVLLHNTQLDNERLNILLP 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3516	98.8	783	4	Q8IWX1	Q8iwx1 homo sapien
2	3512	98.7	783	4	Q8TCS8	Q8tcs8 homo sapien
3	3280	92.2	783	11	Q8K1R3	Q8k1r3 mus musculu
4	2484	69.8	589	11	Q8R2U3	Q8r2u3 mus musculu
5	2389	67.2	540	11	Q9DC52	Q9dc52 mus musculu
6	2285	64.2	504	4	Q96T05	Q96t05 homo sapien
7	1958	55.0	771	5	Q9V9X7	Q9v9x7 drosophila
8	1940.5	54.6	720	5	Q95RX7	Q95rx7 drosophila
9	1761	49.5	717	5	Q8IH29	Q8ih29 drosophila
10	1365.5	38.4	991	10	Q9S7G6	Q9s7g6 arabidopsis
11	1276	35.9	745	5	Q965N3	Q965n3 caenorhabdi
12	1253.5	35.2	707	16	Q8RA43	Q8ra43 thermoanaer
13	1204.5	33.9	703	16	Q97I45	Q97i45 clostridium
14	1198	33.7	733	16	Q8KBY3	Q8kby3 chlorobium
15	1190.5	33.5	698	16	Q99UJ8	Q99uj8 staphylococ
16	1188.5	33.4	698	16	Q8NWY9	Q8nwy9 staphylococ
17	1187.5	33.4	713	2	Q9ZAE1	Q9zae1 thermus the
18	1176	33.1	717	16	Q92SW0	Q92sw0 rhizobium m
19	1170	32.9	703	16	Q8RIA1	Q8ria1 fusobacteri
20	1169	32.9	715	16	Q8DGW9	Q8dgw9 synechococc
21	1166.5	32.8	713	16	Q8UJ56	Q8uj56 agrobacteri
22	1165.5	32.8	701	16	Q8CST1	Q8cst1 staphylococ
23	1163	32.7	717	16	Q8XXP6	Q8xxp6 ralstonia s
24	1161	32.6	711	16	Q8Z3I0	Q8z3i0 salmonella
25	1160.5	32.6	702	16	Q8XJS4	Q8xjs4 clostridium
26	1160	32.6	708	16	Q8DBU9	Q8dbu9 vibrio vuln
27	1160	32.6	711	16	Q8ZLT3	Q8zlt3 salmonella
28	1160	32.6	721	2	Q93MR3	Q93mr3 salmonella
29	1158.5	32.6	749	16	Q92HV7	Q92hv7 rickettsia
30	1158	32.6	734	16	Q8X9M3	Q8x9m3 escherichia
31	1156	32.5	810	16	Q9RSR1	Q9rsr1 deinococcus
32	1153	32.4	718	16	P72659	P72659 synechocyst
33	1153	32.4	740	16	Q8FD87	Q8fd87 escherichia
34	1151	32.4	714	16	Q8FXS9	Q8fxs9 brucella su
35	1150	32.3	714	16	Q8YEB7	Q8yeb7 brucella me
36	1146	32.2	705	16	Q8ZBC6	Q8zbc6 yersinia pe
37	1146	32.2	737	16	Q8D1D1	Q8d1d1 yersinia pe
38	1145	32.2	704	16	Q9KA83	Q9ka83 bacillus ha
39	1140.5	32.1	775	16	O66593	O66593 aquifex aeo
40	1138.5	32.0	718	16	Q8YP11	Q8yp11 anabaena sp
41	1137	32.0	719	16	Q9PGQ9	Q9pgq9 xylella fas
42	1132	31.8	702	16	Q8EHL1	Q8ehl1 shewanella
43	1130	31.8	723	16	Q92C23	Q92c23 listeria in
44	1127	31.7	723	16	Q8Y7F1	Q8y7f1 listeria mo
45	1126	31.7	715	16	Q98BI3	Q98bi3 rhizobium l

ALIGNMENTS

RESULT 1
Q8IWX1
ID Q8IWX1 PRELIMINARY; PRT; 783 AA.
AC Q8IWX1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 3'-5' RNA exonuclease.
GN OLD35.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12473748;
RA Leszczyniecka M., Kang D.-C., Sarkar D., Su Z.-Z., Holmes M.,
RA Valerie K., Fisher P.B.;
RT "Identification and cloning of human polynucleotide phosphorylase,
RT hPNPase (old-35), in the context of terminal differentiation and
RT cellular senescence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16636-16641(2002).
DR EMBL; AY027528; AAK13047.1; -.
KW Exonuclease.
SQ SEQUENCE 783 AA; 85964 MW; 912939AF55309E2C CRC64;

Query Match 98.8%; Score 3516; DB 4; Length 783;
Best Local Similarity 99.4%; Pred. No. 2.9e-219;
Matches 696; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 77

Qy         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 137

Qy        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 197

Qy        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 257

Qy        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 317

Qy        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 377

Qy        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 437

Qy        421 GKVTGLNRRELGHGALAEKALYPVIPRDFPPTIRVTSEVLESNGSSSMASACGGSLALMD 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 GKVTGLNRRELGHGALAEKALYPVIPRDFPPTIRVTSEVLESNGSSSMASACGGSLALMD 497

Qy        481 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 557

Qy        541 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 617

Qy        601 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 677

Qy        661 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
                 |||||||||||||||||||||||||||||||||||  ::
Db        678 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDQRKI 717


RESULT 2
Q8TCS8
ID   Q8TCS8      PRELIMINARY;      PRT;   783 AA.
AC   Q8TCS8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Polynucleotide phosphorylase-like protein (EC 2.7.7.8).
GN   PNPASE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Raijmakers R., Vree Egberts W., Van Venrooij W., Pruijn G.;
RT   "Protein-protein interactions between human exosome components suggest
RT   the assembly of RNase PH-type subunits into a six-membered PNPase-like
RT   ring.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ458465; CAD30289.1; -.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR001547; Glyco_hydro_5.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Nucleotidyltransferase.
SQ   SEQUENCE   783 AA;  85936 MW;  8A3629AF552F8E24 CRC64;

  Query Match           98.7%;  Score 3512;  DB 4;  Length 783;
  Best Local Similarity  99.3%;  Pred. No. 5.3e-219;
  Matches  695;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 77

Qy         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 137

Qy        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 197

Qy        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 257

Qy        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 317

Qy        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 377

Qy        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 437

Qy        421 GKVTGLNRRELGHGALAEKALYPVIPRDFPPTIRVTSEVLESNGSSSMASACGGSLALMD 480
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 GKVTGLNRRELGHGALAEKALYPVIPRDFPPTIRVTSEVLESNGSSSMASACGGSLALMD 497

Qy        481 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 540
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 SGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITAL 557

Qy        541 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 600
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 QADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKR 617

Qy        601 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLE 660
                 |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        618 AKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSAMHEARDFITEICKDDQEQQLE 677

Qy        661 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
                 |||||||||||||||||||||||||||||||||||  ::
Db        678 FGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDQRKI 717


RESULT 3
Q8K1R3
ID   Q8K1R3      PRELIMINARY;      PRT;   783 AA.
AC   Q8K1R3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Polynucleotide phosphorylase-like protein (EC 2.7.7.8).
GN   1200003F12RIK OR PNPASE.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=divers;
RA   Raijmakers R., Vree Egberts W., van Venrooij W., Pruijn G.J.M.;
RT   "Protein-protein interactions between human exosome components support
RT   the assembly of RNase PH-type subunits into a six-membered PNPase-like
RT   ring.";
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ507387; CAD45436.1; -.
DR   MGD; MGI:1918951; 1200003F12Rik.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR001547; Glyco_hydro_5.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Nucleotidyltransferase.
SQ   SEQUENCE   783 AA;  85682 MW;  F35F6BE91AAB5626 CRC64;

  Query Match             92.2%;  Score 3280;  DB 11;  Length 783;
  Best Local Similarity   91.0%;  Pred. No. 5.8e-204;
  Matches  636;  Conservative   39;  Mismatches   24;  Indels    0;  Gaps    0;

Qy          2 GPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDT 61
              ||   | |:|||: ||:|||||| ||||| ||||:||||||||||||||| ||:|||||
Db         19 GPLGRPGRNRALSYLQMRALWSSTGSRAVTVDLGHRKLEISSGKLARFADGCAVIQSGDT 78

Qy         62 AVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSI 121
              |||||||||| ||||||||||||||||||||||||||||||||:|:||:|:||||:|||||
Db         79 AVMVTAVSKTKASPSQFMPLVVDYRQKAAAAGRIPTNYLRREIGSSDREVLTSRVIDRSI 138

Qy        122 RPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDG 181
              ||||||||||||||||||||||:||||:||:|||||||||||||||||||||||||:|||
Db        139 RPLFPAGYFYDTQVLCNLLAVDGINEPDILAVNGASVALSLSDIPWNGPVGAVRIGMIDG 198

Qy        182 EYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQG 241
              | ||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 ECVVNPTRREMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQG 258

Qy        242 IQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRL 301
              |||||| || |||||:|||| ||||||    :|||:||||||||||||||||||||||||
Db        259 IQQLVKEIGVAKRTPQKIFTPSAEIVKYTKIIAMEKLYAVFTDYEHDKVSRDEAVNKIRL 318

Qy        302 DTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFK 361
              |||| |||||||  |  :|||||||:|||||||||:||||||||||||||||||:||||||
Db        319 DTEEHLKEKFPEVDQFEIIESFNIVAKEVFRSIILNEYKRCDGRDLTSLRNISCEVDMFK 378

Qy        362 TLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIG 421
              |||||||||||||||||||||||||| |||||:||||||:|||||||||||||||||||| |
Db        379 TLHGSALFQRGQTQVLCTVTFDSLESSIKSDQIITAINGVKDKNFMLHYEFPPYATNETG 438

Qy        422 KVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMDS 481
              ||||:|||||||||||||||| ||||:|||||||||||||||||||||||||||||||||:
Db        439 KVTGVNRRELGHGALAEKALCPVIPKDFPFTIRVTSEVLESNGSSSMASACGGSLALMDA 498

Qy        482 GVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQ 541
              ||||||||||||:||||||:|||||||||||||||||||||||||||||||||||||||||
Db        499 GVPISSAVAGVAVGLVTKTNPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQ 558

Qy        542 ADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRA 601
              |||||||:|||||:||||||||||||||||||||||||||||||||||||||||:|||||||
Db        559 ADIKLPGVPIKIIMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVKVPLSKRA 618

Qy        602 KFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEF 661
              |||||||||:||||||||||||||||||||||||:||||||:  ||||||||||:||||||||
Db        619 KFVGPGGYHLKKLQAETGVTISQVDEETFSIFAPTPTAMHEARDFITEICRDDQEQQLEF 678

Qy        662 GAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
              ||||||||||||||||||||||||||||||:|||   ::
Db        679 GAVYTATITEIRDTGVMVKLYPNMTAVLLHNSQLDQRKI 717

RESULT 4
Q8R2U3
ID   Q8R2U3      PRELIMINARY;      PRT;   589 AA.
AC   Q8R2U3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to putative.
GN   1200003F12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC027228; AAH27228.1; -.
DR   MGD; MGI:1918951; 1200003F12Rik.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 1.
DR   Pfam; PF03725; RNase_PH_C; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
SQ   SEQUENCE   589 AA;  64962 MW;  808162BB46F6428F CRC64;

  Query Match             69.8%;  Score 2484;  DB 11;  Length 589;
  Best Local Similarity   92.2%;  Pred. No. 1.4e-152;
  Matches  482;  Conservative   26;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        178 IIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQ 237
              :|||| ||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MIDGECVVNPTRREMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQ 60

Qy        238 IIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVN 297
              |||||||||| || |||||:|||| ||||||    :|||:||||||||||||||||||||
Db         61 IIQGIQQLVKEIGVAKRTPQKIFTPSAEIVKYTKIIAMEKLYAVFTDYEHDKVSRDEAVN 120

Qy        298 KIRLDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEV 357
              |||||||| |||||||  |  :|||||||:|||||||||:||||||||||||||||||:||||
Db        121 KIRLDTEEHLKEKFPEVDQFEIIESFNIVAKEVFRSIILNEYKRCDGRDLTSLRNISCEV 180

Qy        358 DMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYAT 417
              ||||||||||||||||||||||||||||||| |||||:||||||:||||||||||||||||
Db        181 DMFKTLHGSALFQRGQTQVLCTVTFDSLESSIKSDQIITAINGVKDKNFMLHYEFPPYAT 240

Qy        418 NEIGKVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLA 477
              || ||||:|||||||||||||||||| ||||:||||||||||||||||||||||||||||||
Db        241 NETGKVTGVNRRELGHGALAEKALCPVIPKDFPFTIRVTSEVLESNGSSSMASACGGSLA 300

Qy        478 LMDSGVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGI 537
              |||:||||||||||||:||||||:||||||||||||||||||||||||||||||||||||||
Db        301 LMDAGVPISSAVAGVAVGLVTKTNPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGI 360

Qy        538 TALQADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPL 597
              |||||||||||:|||||:||||||||||||||||||||||||||||||||||||||||:|||
Db        361 TALQADIKLPGVPIKIIMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVKVPL 420

Qy        598 SKRAKFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQ 657
              |||||||||||||:||||||||||||||||||||||||:||||||:  ||||||||||:||||
Db        421 SKRAKFVGPGGYHLKKLQAETGVTISQVDEETFSIFAPTPTAMHEARDFITEICRDDQEQ 480
```

```
Qy        658 QLEFGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
              ||||||||||||||||||||||||||||||||||||:|||  ::
Db        481 QLEFGAVYTATITEIRDTGVMVKLYPNMTAVLLHNSQLDQRKI 523


RESULT 5
Q9DC52
ID   Q9DC52      PRELIMINARY;      PRT;   540 AA.
AC   Q9DC52;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   1200003F12Rik protein.
GN   1200003F12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Lung;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK004563; BAB23374.1; -.
DR   MGD; MGI:1918951; 1200003F12Rik.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR001547; Glyco_hydro_5.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 1.
DR   PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
SQ   SEQUENCE   540 AA;  58938 MW;  457BFA3E3579A072 CRC64;

  Query Match             67.2%;  Score 2389;  DB 11;  Length 540;
  Best Local Similarity   90.1%;  Pred. No. 1.8e-146;
  Matches  465;  Conservative   29;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          2 GPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDT 61
              ||   | |:|||: ||:|||||| ||||| ||||:|||||||||||||| ||:|||||
Db         19 GPLGRPGRNRALSYLQMRALWSSTGSRAVTVDLGHRKLEISSGKLARFADGCAVIQSGDT 78

Qy         62 AVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSI 121
              |||||||||| |||||||||||||||||||||||||||:|:||:|:|||:|||||
Db         79 AVMVTAVSKTKASPSQFMPLVVDYRQKAAAAGRIPTNYLRREIGSSDREVLTSRVIDRSI 138

Qy        122 RPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDG 181
              |||||||||||||||||||||:||||:||:|||||||||||||||||||||||:|||
Db        139 RPLFPAGYFYDTQVLCNLLAVDGINEPDILAVNGASVALSLSDIPWNGPVGAVRIGMIDG 198

Qy        182 EYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQG 241
              | |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 ECVVNPTRREMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQG 258

Qy        242 IQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRL 301
              ||||||| || ||||||:|||| ||||||  :|||:|||||||||||||||||||||||
Db        259 IQQLVKEIGVAKRTPQKIFTPSAEIVKYTKIIAMEKLYAVFTDYEHDKVSRDEAVNKIRL 318

Qy        302 DTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFK 361
              |||| ||||||| | :|||||||:|||||||:|||||||||||||||:||||||||
Db        319 DTEEHLKEKFPEVDQFEIIESFNIVAKEVFRSIILNEYKRCDGRDLTSLRNISCEVDMFK 378

Qy        362 TLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIG 421
              |||||||||||||||||||||||||| |||||:|||||:|||||||||||| ||||| |
Db        379 TLHGSALFQRGQTQVLCTVTFDSLESSIKSDQIITAINGVKDKNFMLHYEFPLYATNETG 438

Qy        422 KVTGLNRRELGHGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMDS 481
              ||||:|||||||||||||| ||||:||||||||||||||||||||||||||||||||:
Db        439 KVTGVNRRELGHGALAEKALCPVIPKDFPFTIRVTSEVLESNGSSSMASACGGSLALMDA 498

Qy        482 GVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDIL 517
              ||||||||||||:|||||||||||||||||||
Db        499 GVPISSAVAGVAVGLVTKTNPEKGEIEDYRLLTDIL 534


RESULT 6
Q96T05
ID   Q96T05      PRELIMINARY;      PRT;   504 AA.
AC   Q96T05;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein FLJ14531.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK027437; BAB55109.1; -.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR001547; Glyco_hydro_5.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 1.
DR   PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   504 AA;  55998 MW;  2BB89ADB409322D6 CRC64;

  Query Match             64.2%;  Score 2285;  DB 4;  Length 504;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-140;
  Matches  450;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 DGPFLLPRRDRALTQLQVRALWSSAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGD 77

Qy         61 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 TAVMVTAVSKTKPSPSQFMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRS 137

Qy        121 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 IRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIID 197

Qy        181 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 GEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQ 257
```

```
Qy        241 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 GIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIR 317

Qy        301 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 LDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMF 377

Qy        361 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 KTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEI 437

Qy        421 GKVTGLNRRELGHGALAEKALYPVIPRDFP 450
              |||||||||||||||||||||||||||||
Db        438 GKVTGLNRRELGHGALAEKALYPVIPRDFP 467
```

```
RESULT 7
Q9V9X7
ID   Q9V9X7      PRELIMINARY;      PRT;   771 AA.
AC   Q9V9X7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG11337 protein.
GN   CG11337.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003778; AAF57151.2; -.
DR   HSSP; P05055; 1SRO.
DR   FlyBase; FBgn0039846; CG11337.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
SQ   SEQUENCE   771 AA;  84953 MW;  779F086DB17022C4 CRC64;
```

```
  Query Match             55.0%;  Score 1958;  DB 5;  Length 771;
  Best Local Similarity   55.2%;  Pred. No. 2.7e-118;
  Matches  380;  Conservative  128;  Mismatches  173;  Indels    8;  Gaps    5;

Qy         19 RALWSSAGSRAVAVDL---GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSP 75
              | : ||:   | :|::     | :  |||:|||||:|:|| | ||||||||||:| ||:|
Db         29 RGIQSSSNGEAPSVEVNPSNGRNMTFSSGRLARFANGTAVCQMGDTAVMVTAVAKAKPNP 88

Qy         76 SQ-FMPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQ 134
              | |||||||||| | ||:|||| |::|||:| |:||||::|:||||:|||| | :||
Db         89 GQGFMPLVVDYRLKNAASGRIPMNFMRRELGPSEKEILSARLIDRSLRPLFHKDYRTETQ 148

Qy        135 VLCNLLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSS 194
              ::||:||:| |: ||||||| ||:|||||||||||:||||:|: ||| ::||||:|: :
Db        149 LVCNMLAMDAVHSPDVLAINAASMALSLSDIPWNGPIGAVRVGLCDGEVLINPTRRELQT 208

Qy        195 STLNLVVAGAPKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKR 254
              | |:|||:   :: :|||     :||||   ||| | :  | ||  |::| |  |  ||
Db        209 SQLDLVVSATKQNLVVMLEGKGNVVLQQDLLKAIKQGTREAQFIIHEIERLQKAYGRQKR 268

Qy        255 TPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEA 314
                :     ||: |    :   ||  :| |  |||:||| |||::| :  :::   ||:
Db        269 EVEVAAEVDPELGKAVRSMCEMRLREIFQDSTHDKMSRDNAVNEVRSNVIDKVWSSFPDT 328
```

```
Qy        315 DPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQT 374
Db        329 EPSLITEQFNQTSRTIFRELIFERGLRCDGRDYDQLRNISCQVDMYKPLHGSALFQRGQT 388

Qy        375 QVLCTVTFDSLESGIKSDQVITAIN--GIKDKNFMLHYEFPPYATNEIGKVTGLNRRELG 432
Db        389 QVFCTVSLDSQESAMKLDS-LAALDSGGLKAKNFMLHYEFPPYATGEVGRIGPVGRREMG 447

Qy        433 HGALAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGV 492
Db        448 HGALAERSLLPTLPNDYPFTVRLTSEVLESNGSSSMASVCGGSLALMDAGVPVSAPAAGV 507

Qy        493 AIGLVTK-TDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPI 551
Db        508 AIGLVTKFENDDTKHLQDYRILTDILGIEDYMGDMDMKVAGTRKGFTAIQADLKIPGIPL 567

Qy        552 KIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNL 611
Db        568 KVVMESLQKATDAKSNILDIMSEAIREPRKYPKESWPVSETLTVEPQQRAQLIGPSGLHM 627

Qy        612 KKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITE 671
Db        628 KRIYLETGTSLTAVDETHFNVFAPSQAAMDEAKELIEGYMVKERVPDLEFGGIYTAKITE 687

Qy        672 IRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
Db        688 LRDTGVMVILYPSMPPALLHNSQLDQRKI 716


RESULT 8
Q95RX7
ID   Q95RX7      PRELIMINARY;      PRT;   720 AA.
AC   Q95RX7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   LD03255p.
GN   CG11337.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY061061; AAL28609.1; -.
DR   FlyBase; FBgn0039846; CG11337.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
SQ   SEQUENCE   720 AA;  79382 MW;  DD54952528E5DBDF CRC64;

  Query Match           54.6%;  Score 1940.5;  DB 5;  Length 720;
  Best Local Similarity  56.2%;  Pred. No. 3.3e-117;
  Matches  374;  Conservative  123;  Mismatches  164;  Indels    5;  Gaps    4;

Qy         39 LEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQ-FMPLVVDYRQKAAAAGRIPT 97
Db          1 MTFSSGRLARFANGTAVCQMGDTAVMVTAVAKAKPNPGQGFMPLVVDYRLKNAASGRIPM 60

Qy         98 NYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVLAINGAS 157
Db         61 NFMRRELGPSEKEILSARLIDRSLRPLFHKDYRTETQLVCNMLAMDAVHSPDVLAINAAS 120

Qy        158 VALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVMLEASAE 217
Db        121 MALSLSDIPWNGPIGAVRVGLCDGEVLINPTRRELQTSQLDLVVSATKQNLVVMLEGKGN 180

Qy        218 NILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTHKLAMER 277
Db        181 VVLQQDLLKAIKQGTREAQFIIHEIERLQKAYGRQKREVEVAAEVDPELGKAVRSMCEMR 240

Qy        278 LYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAKEVFRSIVLN 337
Db        241 LREIFQDSTHDKMSRDNAVNEVRSNVIDKVWSSFPDTEPSLITEQFNQTSRTIFRELIFE 300

Qy        338 EYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKSDQVITA 397
Db        301 RGLRCDGRDYDQLRNISCQVDMYKPLHGSALFQRGQTQVFCTVSLDSQESAMKLDS-LAA 359

Qy        398 IN--GIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIPRDFPFTIRV 455
Db        360 LDSGGLKAKNFMLHYEFPPYATGEVGRIGPVGRREMGHGALAERSLLPTLPNDYPFTVRL 419

Qy        456 TSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTK-TDPEKGEIEDYRLLT 514
Db        420 TSEVLESNGSSSMASVCGGSLALMDAGVPVSAPAAGVAIGLVTKFENDDTKHLQDYRILT 479

Qy        515 DILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKEILQIMNK 574
Db        480 DILGIEDYMGDMDMKVAGTRKGFTAIQADLKIPGIPLKVVMESLQKATDAKSNILDIMSE 539

Qy        575 TISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDEETFSVFA 634
Db        540 AIREPRKYPKESWPVSETLTVEPQQRAQLIGPSGLHMKRIYLETGTSLTAVDETHFNVFA 599

Qy        635 PTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQ 694
Db        600 PSQAAMDEAKELIEGYMVKERVPDLEFGGIYTAKITELRDTGVMVILYPSMPPALLHNSQ 659

Qy        695 LDNERL 700
Db        660 LDQRKI 665


RESULT 9
Q8IH29
ID   Q8IH29      PRELIMINARY;      PRT;   717 AA.
AC   Q8IH29;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   GM16802p (Fragment).
GN   CG11337.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BT001457; AAN71212.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   717 AA;  79404 MW;  6C1CA7ABC1714D02 CRC64;
```

```
  Query Match             49.5%;  Score 1761;  DB 5;  Length 717;
  Best Local Similarity   54.4%;  Pred. No. 1.4e-105;
  Matches  336;  Conservative  119;  Mismatches  159;  Indels    4;  Gaps    3;

Qy         86 RQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGV 145
Db         46 RNMTFSSGRIPMNFMRRELGPSEKEILSARLIDRSLRPLFHKDYRTETQLVCNMLAMDAV 105

Qy        146 NEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAP 205
Db        106 HSPDVLAINAASMALSLSDIPWNGPIGAVRVGLCDGEVLINPTRRELQTSQLDLVVSATK 165

Qy        206 KSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPE 265
Db        166 QNLVVMLEGKGNVVLQQDLLKAIKQGTREAQFIIHEIERLQKAYGRQKREVEVAAEVDPE 225

Qy        266 IVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNV 325
Db        226 LGKAVRSMCEMRLREIFQDSTHDKMSRDNAVNEVRSNVIDKVWSSFPDTEPSLITEQFNQ 285

Qy        326 VAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSL 385
Db        286 TSRTIFRELIFERGLRCDGRDYDQLRNISCQVDMYKPLHGSALFQRGQTQVFCTVSLDSQ 345

Qy        386 ESGIKSDQVITAIN--GIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYP 443
Db        346 ESAMKLDS-LAALDSGGLKAKNFMLHYEFPPYATGEVGRIGPVGRREMGHGALAERSLLP 404

Qy        444 VIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTK-TDP 502
Db        405 TLPNDYPFTVRLTSEVLESNGSSSMASVCGGSLALMDAGVPVSAPAAGVAIGLVTKFEND 464

Qy        503 EKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQAS 562
Db        465 DTKHLQDYRILTDILGIEDYMGDMDMKVAGTRKGFTAIQADLKIPGIPLKVVMESLQKAT 524

Qy        563 VAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTI 622
Db        525 DAKSKILDIMSEAIREPRKYPKESWPVSETLTVEPQQRAQLIGPSGLHMKRIYLETGTSL 584

Qy        623 SQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLY 682
Db        585 TAVDETHFNVFAPSQAAMDEAKELIEGYMVKERVPDLEFGGIYTAKITELRDTGVMVILY 644

Qy        683 PNMTAVLLHNTQLDNERL 700
Db        645 PSMPPALLHNSQLDQRKI 662


RESULT 10
Q9S7G6
ID   Q9S7G6      PRELIMINARY;      PRT;   991 AA.
AC   Q9S7G6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Polynucleotide phosphorylase.
GN   PNP OR T15N1_70.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Kim Y.J., Zhou D.X., Mache R.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA   Rudd S., Lemcke K., Mayer K.F.X.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y14686; CAB43865.1; -.
DR   EMBL; Y14685; CAB43864.1; -.
DR   EMBL; AL163792; CAB87625.1; -.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 2.
DR   PROSITE; PS50126; S1; 1.
KW   Transit peptide.
SQ   SEQUENCE   991 AA;  107771 MW;  9557E25980C6D1A5 CRC64;

  Query Match             38.4%;  Score 1365.5;  DB 10;  Length 991;
  Best Local Similarity   41.7%;  Pred. No. 1e-79;
  Matches  287;  Conservative  138;  Mismatches  231;  Indels   33;  Gaps   11;

Qy         24 SAGSRAV-----AVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQF 78
Db         44 SAGTKILESFKEEFEVGSRVVSFETGKIARFANGSVVLGMDETKVLSTVTCAKTDSPRDF 103

Qy         79 MPLVVDYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCN 138
Db        104 LPLTVDYQEKQYAQGLIPNTYMRREGAPKERELLCGRLIDRPIRPLFPTGFYHEVQIMAS 163

Qy        139 LLAVDGVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLN 198
Db        164 VLSSDGKQDPDILAANASSAALMLSDVPWGGPIGVIRIGRICGQFVVNPTMDELSSSDLN 223

Qy        199 LVVAGAPKSQIVMLEASAENILQQDFCHAIKV----GVKYTQQIIQGIQQLVKETGVTKR 254
Db        224 LIYA-CTRDKTMMIDVQSREISEKDLAAALRLAHPEAVKYLDPQI----RLAEKAGKQKK 278

Qy        255 TPQKLFTPSPEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEA 314
Db        279 -EYKLSMLSDKTLEKVADLAATRIESVFTDPSYGKFERGEALDNIGKDVRKVFEEEGDQE 337

Qy        315 DPYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQT 374
Db        338 SLSILPKAVDTVRKKVVRSRMISDGFRVDGRHVDEVRPIYCESHYLPALHGSALFSRGDT 397

Qy        375 QVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHG 434
Db        398 QVLCTVTLGAPAEAQSLDSLV----GPPKKRFMLHYSFPPYCTNEVGKRGGLNRREVGHG 453

Qy        435 ALAEKALYPVIPRD--FPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGV 492
Db        454 TLAEKALLAVLPPEEAFPYTIRINSEVMSSDGSTSMASVCGGSMALMDAGIPLRAHVAGV 513

Qy        493 AIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIK 552
Db        514 SVGLITDVDPSSGEIKDYRIVTDILGLEDHLGDMDFKIAGTRDGVTAIQLDIKPAGIPLD 573

Qy        553 IVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLK 612
Db        574 IVCESLENAREARLQILDHMERNINSPRGQDGAYSPRLATLKYSNDSLRTLIGPMGVLKR 633

Qy        613 KLQAETGVTISQVDEETFSVFAPTPSVMHEAR---DFITEICKDDQEQQLEFGAVYTATI 669
Db        634 KIEVETGARLS-IDNGTLTIVAKNQDVMEKAQEQVDFII-------GRELVVGGVYKGTV 685

Qy        670 TEIRDTGVMVKLYPNMTAVLLHNTQLDNE 698
Db        686 SSIKEYGAFVE-FPGGQQGLLHMSELSHE 713
```

```
RESULT 11
Q965N3
ID   Q965N3      PRELIMINARY;      PRT;   745 AA.
AC   Q965N3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 82.9 kDa protein.
GN   BE0003N10.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC092690; AAK73855.2; -.
DR   WormPep; BE0003N10.1; CE30235.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   745 AA;  82881 MW;  45D03070C72ACB16 CRC64;

  Query Match           35.9%;  Score 1276;  DB 5;  Length 745;
  Best Local Similarity  42.2%;  Pred. No. 4.1e-74;
  Matches  278;  Conservative  109;  Mismatches  237;  Indels   34;  Gaps   10;

Qy         28 RAVAVDL-GNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVS-KTKPSPSQFMPLVVDY 85
              ::|||| |   :|: :| ||||| : |   || ||| | |  |:|      :|| |:|
Db         26 QSVAVDFDGKSGIELKAGHLARFASSTVVASFGDNAVMATVVQRKSKQGDGSGVPLQVEY 85

Qy         86 RQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGV 145
              | |:| ||||||:||||:  || ||| || |||||||| | |  |:||::|  ||:|
Db         86 RPSASAIGRIPTNFLRRELSQSDNEILISRAIDRSIRPLLP-GNSYETQIICKPLALDEN 144

Qy        146 NEPDVLAINGASVALSLSDIPWNGPVGAVRIG-IIDGEYVVNPTRKEMSSSTLNLVVAGA 204
             :  :| || || || :|   :|||: |:|:     |:: ||||::|:  ::|||:||
Db        145 ADQIMLGINAASTALQISSAAYNGPLAALRVARTARGDFHVNPTQEELREASLNLIVAMR 204

Qy        205 PKSQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSP 264
             : ||:|         :   ||: |  ::  :: : ::||  |              |
Db        205 KHEKTVMIELDGRESSAEHLEHALDVAFRHVAKLHEAMEQLTAE------------PKD 251

Qy        265 EIV--------KYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEK-FPEAD 315
             |:         :   : | ||:| | ||  |||:|||  :  |    ||   || |   :
Db        252 ELASEDFSGLERLLEETARERIYYVITDAGHDKISRDMEIKAI---FEEICAEKAFQTCE 308

Qy        316 PYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQ 375
             |  ::: : |:| |   |    |||||      | ::  |||:| ||| ::|||||||
Db        309 KDAIYRTYSTLVKKVLRDTTLRTGIRCDGRRPDEFRPITIHVDMYKKLHGCSIFQRGQTQ 368

Qy        376 VLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGA 435
             |: ||||||  :    | |   :   : |:||||||| :| || |    |||||:||||
Db        369 VMSTVTFDSPAAAFHPDSVAQLLGSQRKKSFMLHYEFPGFAINEFGTTRSLNRREIGHGA 428

Qy        436 LAEKALYPVIPRDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIG 495
             ||||:|  : | |||:  |:  :|||||||||||| ||||||| |:|||: :| ||||||
Db        429 LAEKSLKNLFPADFPYATRLACQVLESNGSSSMASVCGGSLALFDAGVPMKNAAAGVAIG 488

Qy        496 LVT-KTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIV 554
             |:: : :||      ||:|||||||||| |||||:|||  | || | |:|  |:  : :
Db        489 LISDEAEPE----TKYRVLTDILGIEDYAGDMDFKVAGTRDGFTAAQLDVKNGGLTRRQL 544

Qy        555 MEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKL 614
             |::| |      :|| |:    :||   |   |:::::::   ||       ||| | :
Db        545 TESLQAARAGIDHVLQKMSVMRDRPREQFKPTVPIIQSMRIEPHKRVTLFRNNGYNCKLI 604

Qy        615 QAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEI 672
             :||||| ||  ||   |: |     : :| | : :: : :      ||::  | | |:
Db        605 EAETGVKISAEDEAHISLLAQDKEKLQKAMDMMNDVLESNSTLDFAFGSIVQAEIVEV 662


RESULT 12
Q8RA43
ID   Q8RA43      PRELIMINARY;      PRT;   707 AA.
AC   Q8RA43;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Polyribonucleotide nucleotidyltransferase (Polynucleotide
DE   phosphorylase).
GN   PNP OR TTE1387.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC   Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4 / JCM 11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013098; AAM24609.1; -.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   707 AA;  78519 MW;  779E64C04BBA3331 CRC64;

  Query Match           35.2%;  Score 1253.5;  DB 16;  Length 707;
  Best Local Similarity  40.0%;  Pred. No. 1.1e-72;
  Matches  275;  Conservative  135;  Mismatches  230;  Indels   47;  Gaps   13;

Qy         28 RAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPS-QFMPLVVDYR 86
              |   ::|  ||| : :||:|: |:|:| |: ||| |:||| :  :|      | || |:|
Db          4 RTFEMELAGRKLVVQTGKVAQQANGAAWVKYGDTVVLVTACASKEPREGIDFFPLTVEYE 63

Qy         87 QKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVN 146
              ::  : |:||  :::||    |:| ||::|:||| |||||| ||  | ||:  :|:||
Db         64 ERLYSIGKIPGGFIKREGKPSEKAILSARLIDRPIRPLFPEGYRNDVQVIATVLSVDPDA 123

Qy        147 EPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPK 206
             :|:::|: |:|||||:||||:||| |:| :|::||::|:||| ::    | ::|||:|  |
Db        124 QPEIVAMIGSSVALSISDIPFNGPTGSVAVGLVDGKFVINPTYEQREKSLMHLVVSGT-K 182

Qy        207 SQIVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEI 266
              |||:|| |: : ::    ||    :| :||:: |: :||| ||           |  |:
Db        183 DAIVMVEAGAKEVPEETILDAIVYAHQYIKQIVEFIEGIVKEVGV----------PKAEV 232
```

```
Qy        267 VKYTHKLAME-----RLYAVFTDYE----HDKVSRDEAVNKIRLDTEEQLKEKFPE--AD 315
Db        233 V--LHEIDKELEEKVRAYATEKIYNALRTPEKKERNDNLDKVEQEVLEHFKDEYPDNLAD 290

Qy        316 PYEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQ 375
Db        291 IDEVLYK---IMKEQMRKMIKEEKIRVDGRGLDDIRPIWCEVGVLPRTHGSAIFTRGQTQ 347

Qy        376 VLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGA 435
Db        348 VLTVATL----GAIGDIQILEGIGDEEFKRYMHHYNFPPYSVGEVRPLRGPGRREIGHGA 403

Qy        436 LAEKALYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVA 493
Db        404 LAERALEPVIPSEEEFPYTIRLVSEVLSSNGSTSQASVCGSTLALMDAGVPIKAPVAGIA 463

Qy        494 IGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKI 553
Db        464 MGLIKEED-------EVLILTDIQGIEDFLGDMDFKVAGTEKGVTAIQMDIKIPGIDRDI 516

Qy        554 VMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKK 613
Db        517 LQMALEKARKARLYVLQKMLEVIKEPRKELSKYAPRVVRMVVNPEKIRDIIGPAGKTITK 576

Qy        614 LQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIR 673
Db        577 IISETGVKIDIEEDGRLYITAPNLEAGERAKQMIEAITKD-----IEVGGIYLGKVLRIA 631

Qy        674 DTGVMVKLYPNMTAVLLHNTQLDNERL 700
Db        632 PFGAFVEIAPGKEG-LVHISNLSKKRV 657
```

RESULT 13
Q97I45
ID Q97I45 PRELIMINARY; PRT; 703 AA.
AC Q97I45;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Polyribonucleotide nucleotidyltransferase.
GN CAC1808.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007689; AAK79773.1; -.
DR InterPro; IPR001247; 3_ExoRNase.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR003029; S1.
DR Pfam; PF00013; KH; 1.
DR Pfam; PF03726; PNPase; 1.
DR Pfam; PF01138; RNase_PH; 2.
DR Pfam; PF03725; RNase_PH_C; 2.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00322; KH; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PS50084; KH_TYPE_1; 1.
DR PROSITE; PS50126; S1; 1.
KW Transferase; Complete proteome.

SQ SEQUENCE 703 AA; 77989 MW; 5EABC83EDEDCA22B CRC64;

Query Match 33.9%; Score 1204.5; DB 16; Length 703;
Best Local Similarity 39.6%; Pred. No. 1.6e-69;
Matches 266; Conservative 125; Mismatches 258; Indels 23; Gaps 10;

```
Qy         33 DLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPS-QFMPLVVDYRQKAAA 91
Db          8 DIAGRKLKVECGKTGMLSNCAMFISYGDTVVMVNVNASEKPREGIDFFPLSIEYEERQYS 67

Qy         92 AGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEPDVL 151
Db         68 VGKIPGGFVKREGRPSEKSILHARAIDRPLRPLFPKGYRNDVQVVCTVMSVEQDNLPEIL 127

Qy        152 AINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQIVM 211
Db        128 AMNGASMALCLSDIPFTTPVATVSVGCIDGKFVLNPTLEEREKSSLDLTVC-ATNERVMM 186

Qy        212 LEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPSPEIVKYTH 271
Db        187 LEAGADEIPEDLMIAAIDFGFNACQDIVAFQEKAMKEFGKEKVTPE-LYHPKEEIEKDVT 245

Qy        272 KLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFPEADPYEIIESFNVVAKEVF 331
Db        246 EFAFESIKEIM--YITDRDERNLRLREIKEKISNEFAEKYPD-DGADIDEVVYTLQKKVV 302

Qy        332 RSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGSALFQRGQTQVLCTVTFDSLESGIKS 391
Db        303 RNMLLKEHRRPDGRRFDEIRPISCDVDLLPRTHGSGLFTRGLTQVMTVTTL----GPIGD 358

Qy        392 DQVITAINGIKDKNFMLHYEFPPYATNEIGKVTGLNRRELGHGALAEKALYPVIP--RDF 449
Db        359 AQVIDGLGVEESKRYMHHYNFPPYSTGEVKPLRGPNRREIGHGALAEKALVPLIPSEEEF 418

Qy        450 PFTIRVTSEVLESNGSSSMASACGGSLALMDSGVPISSAVAGVAIGLVTKTDPEKGEIED 509
Db        419 PYTIRLVSEVLSSNGSTSQASVCGSTLALMDAGVPIKRPAAGIAMGLITSEDLSKEAV-- 476

Qy        510 YRLLTDILGIEDYNGDMDFKIAGTNKGITALQADIKLPGIPIKIVMEAIQQASVAKKEIL 569
Db        477 ---ITDIQGLEDFFGDMDFKVAGTEKGITAIQVDTKIHGLSKYCIKTAINDARKARLFIL 533

Qy        570 QIMNKTISKPRASRKENGPVVETVQVPLSKRAKFVGPGGYNLKKLQAETGVTISQVDEET 629
Db        534 EKMVACINEPRKELSTYAPRAYTINIDTDKIRTLIGTGGKTINKIIEETGVKIDIREDGT 593

Qy        630 FSVFAPTPSVMHEARDFITEICKDDQEQQLEFGAVYTATITEIRDTGVMVKLYPNMTAVL 689
Db        594 VFVLSSDADSANRALKMIDDLTKD-----VKVGEVYLGKVTKITNFGAFVEVLPGKEG-L 647

Qy        690 LHNTQLDNERLN 701
Db        648 VHISKLDINKVN 659
```

RESULT 14
Q8KBY3
ID Q8KBY3 PRELIMINARY; PRT; 733 AA.
AC Q8KBY3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Polyribonucleotide nucleotidyltransferase.
GN PNP OR CT1649.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,

```
RA   Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA   Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA   Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA   Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA   Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA   Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT   "The complete genome sequence of Chlorobium tepidum TLS, a
RT   photosynthetic, anaerobic, green-sulfur bacterium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR   EMBL; AE012920; AAM72874.1; -.
DR   TIGR; CT1649; -.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   733 AA;  79937 MW;  3C5D0C71E00C6AD9 CRC64;

  Query Match           33.7%;  Score 1198;  DB 16;  Length 733;
  Best Local Similarity   38.1%;  Pred. No. 4.6e-69;
  Matches  266;  Conservative  144;  Mismatches  232;  Indels   56;  Gaps   16;

Qy         32 VDLGNRK-LEISSGKLARFADGSAVVQSGDTAVMVTAV-SKTKPSPSQ-FMPLVVDYRQK 88
              :|||: | : | :||:|: |||||||   || |: | | ||| |||:| | || |:||:|
Db          7 IDLGHGKVITIETGKMAKQADGSAVVTMNDTMVLATVVSSKTPPSPNQDFFPLQVEYREK 66

Qy         89 AAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVDGVNEP 148
              :|||: |  : :||   |:||||::|:|||::||||| ||: :||:: :::: | :|:
Db         67 YSAAGKFPGGFFKREGRPSEKEILSARLIDRALRPLFPDGYYQETQIIISVISSDTINDA 126

Qy        149 DVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAGAPKSQ 208
              |||    || |: :||||:  |:  ||:| |:| ::|||   |:: | ::: : |  :
Db        127 DVLGGIAASAAIMVSDIPFANPMSEVRVGRINGLFIVNPDINELAQSDMDICIGGT-EDT 185

Qy        209 IVMLEASAENILQQDFCHAIKVGVKYTQQIIQGIQQLVKETGVTKRTPQKLFTPS---PE 265
              | |||   : | : :   |||  |       |: |  | :|       |:: |:|:    |
Db        186 ICMLEGEMKEISEAEMLDAIKFG----HDAIKKICALQRELAAEVAKPKRPFSPTVAPDE 241

Qy        266 IVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKEKFP----------EA 314
              :| :  :     | |:      :  ::::|   | :    : :::            ||
Db        242 LVNFVEEHCSAELKAL----AYTPLAKEERAEKTKAIYTQTIRKTLTHFTDRVGPDQIEA 297

Qy        315 DP--------YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTLHGS 366
              ||        : | |  : | |:| | ::|:: || ||| |  :| :| |: :    |||
Db        298 DPTSAFCLNEHMIEECIHAVEKKVMRHMILDDGKRLDGRTLEQVRPISIELGLIPRAHGS 357

Qy        367 ALFQRGQTQVLCTVTFDSLESGIKSD-QVITAINGIKDKNFMLHYEFPPYATNEIGKVTG 425
              ||| ||:|| | |:|      | | | :  :   ||| ||||| |||::  |||:| |
Db        358 ALFTRGETQALVTLTL-----GTKKDAQSVDTLTDDKDKRFMLHYNFPPFSVGEIGRVGG 412

Qy        426 LNRRELGHGALAEKALYPVIP--RDFPPFTIRVTSEVLESNGSSSMASACGGSLALMDSGV 483
              |||:||| |||:|:  |:|  ::||:|:|: |::|||||||||| |||:|| || |:
Db        413 AGRREIGHGNLAERAIKMVMPSEQEFPYTVRLVSDILESNGSSSMASVCGGTLAAMDGGI 472

Qy        484 PISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQAD 543
              |:   |:|:|:||: : |        | :|:|||| ||: ||||||:||| |||| | |
Db        473 PLKKPVSGIAMGLIKEGD-------RYAVLSDILGNEDHLGDMDFKVAGTRDGITACQMD 525

Qy        544 IKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRAKF 603
              ||: |:    |:  |::||   :  || :| : | : ||   :  | : |:|:|:
Db        526 IKIDGLDYHILETALEQARKGRLHILDVMAEAIPESRADIGKYAPRLTTIQIPVDAIGMV 585

Qy        604 VGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEFGA 663
              :| ||  :: :  |||  |:  |: | ::   :|     | : |  :       : | |
Db        586 IGKGGETIRSITEETGAEINIDDDGTVTIACSSPEATKAAVETIKTLV-----SKPEVGT 640

Qy        664 VYTATITEIRD-TGVMVKLYPNMTAVLLHNTQLDNERL 700
              :|   : :|||  |  |:  |  |  |:| ::: ||:
Db        641 IYMGKVRDIRDELGAFVEFLPK-TDGLVHISEIARERI 677


RESULT 15
Q99UJ8
ID   Q99UJ8      PRELIMINARY;      PRT;   698 AA.
AC   Q99UJ8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Polyribonucleotide nucleotidyltransferase.
GN   PNPA OR SAV1274 OR SA1117.
OS   Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS   Staphylococcus aureus (strain N315).
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=158878, 158879;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
DR   EMBL; AP003361; BAB57436.1; -.
DR   EMBL; AP003133; BAB42369.1; -.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   InterPro; IPR004087; KH_dom.
DR   InterPro; IPR004088; KH_type_1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00013; KH; 1.
DR   Pfam; PF03726; PNPase; 1.
DR   Pfam; PF01138; RNase_PH; 2.
DR   Pfam; PF03725; RNase_PH_C; 2.
DR   Pfam; PF00575; S1; 1.
DR   SMART; SM00322; KH; 1.
DR   SMART; SM00316; S1; 1.
DR   PROSITE; PS50084; KH_TYPE_1; 1.
DR   PROSITE; PS50126; S1; 1.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   698 AA;  77361 MW;  6EABFB86EDFBFB2E CRC64;

  Query Match           33.5%;  Score 1190.5;  DB 16;  Length 698;
  Best Local Similarity   39.1%;  Pred. No. 1.3e-68;
  Matches  273;  Conservative  118;  Mismatches  249;  Indels   59;  Gaps   12;

Qy         24 SAGSRAVAVDLGNRKLEISSGKLARFADGSAVVQSGDTAVMVTAVSKTKPSPSQFMPLVV 83
              |   :    | | | :|:||: |:|: :|: ||| |: || :  :|    | || |
Db          2 SQEKKVFKTEWAGRSLTIETGQLAKQANGAVLVRYGDTVVLSTATASKEPRDGDFFPLTV 61

Qy         84 DYRQKAAAAGRIPTNYLRREVGTSDKEILTSRIIDRSIRPLFPAGYFYDTQVLCNLLAVD 143
              :| :|  |||:||  : :||     |   ||:|:||| |||||| || :| |::  :|: |
Db         62 NYEEKMYAAGKIPGGFKKREGRPGDDATLTARLIDRPIRPLFPKGYKHDVQIMNMVLSAD 121

Qy        144 GVNEPDVLAINGASVALSLSDIPWNGPVGAVRIGIIDGEYVVNPTRKEMSSSTLNLVVAG 203
                | : |: |:|:|||:|||: ||:  | :|  |||:|::||| :|    | |:| |||
Db        122 PDCSPQMAAMIGSSMALSVSDIPFQGPIAGVNVGYIDGKYIINPTVEEKEVSRLDLEVAG 181

Qy        204 APKSQIVMLEASAENILQQDFCHAIKVG-------VKYTQQIIQGIQQLVKETGVTKRTP 256
                |  : |:|| |  | :|:   || |        | : |||:  || : :|
Db        182 -HKDAVNMVEAGASEITEQEMLEAIFFGHEEIQRLVDFQQQIVDHIQPVKQE-------- 232
```

```
Qy        257 QKLFTPS-------PEIVKYTHKLAMERLYAVFTDYEHDKVSRDEAVNKIRLDTEEQLKE 309
              | |:       :  |  :  ::   .  |     ||  ||| :: ::   || : |
Db        233 ---FIPAERDEALVERVKSLTEEKGLKETVLTF-----DKQQRDENLDNLK---EEIVNE 281

Qy        310 KFPEADP------YEIIESFNVVAKEVFRSIVLNEYKRCDGRDLTSLRNVSCEVDMFKTL 363
                | ||       |:    | : ||  | :: :|  | |||    :| :  || :
Db        282 FIDEEDPENELLIKEVYAILNELVKEEVRRLIADEKIRPDGRKPDEIRPLDSEVGILPRT 341

Qy        364 HGSALFQRGQTQVLCTVTFDSLESGIKSDQVITAINGIKDKNFMLHYEFPPYATNEIGKV 423
              ||| || ||||| |  :|  :|       |:|  :   ::|  || || ||  ::  | | |
Db        342 HGSGLFTRGQTQALSVLTLGAL----GDYQLIDGLGPEEEKRFMHHYNFPNFSVGETGPV 397

Qy        424 TGLNRRELGHGALAEKALYPVIP--RDFPFTIRVTSEVLESNGSSSMASACGGSLALMDS 481
                |||:||||| |:||  :||   |||:|||: |||||||||| || || :|||||:
Db        398 RAPGRREIGHGALGERALKYIIPDTADFPYTIRIVSEVLESNGSSSQASICGSTLALMDA 457

Qy        482 GVPISSAVAGVAIGLVTKTDPEKGEIEDYRLLTDILGIEDYNGDMDFKIAGTNKGITALQ 541
              |||| : |||:|:|||| |         | :|||| |:||  ||||||:||| :||||:|
Db        458 GVPIKAPVAGIAMGLVTRED-------SYTILTDIQGMEDALGDMDFKVAGTKEGITAIQ 510

Qy        542 ADIKLPGIPIKIVMEAIQQASVAKKEILQIMNKTISKPRASRKENGPVVETVQVPLSKRA 601
              |||: |:  :|: ||::||   : ||:  | :|| :||        | | |: :   |
Db        511 MDIKIDGLTREIIEEALEQARRGRLEIMNHMLQTIDQPRTELSAYAPKVVTMTIKPDKIR 570

Qy        602 KFVGPGGYNLKKLQAETGVTISQVDEETFSVFAPTPSVMHEARDFITEICKDDQEQQLEF 661
              :||||  : ::  |||| :    : |  : |   :::: ||: | || :     : |
Db        571 DVIGPGGKKINEIIDETGVKLDIEQDGTIFIGAVDQAMINRAREIIEEITR-----EAEV 625

Qy        662 GAVYTATITEIRDTGVMVKLYPNMTAVLLHNTQLDNERL 700
              |  | ||:  |   |  | |:|   | ||| :|:   |:
Db        626 GQTYQATVKRIEKYGAFVGLFPGKDA-LLHISQISKNRI 663
```

Search completed: January 8, 2004, 10:40:18
Job time : 50 secs